

Launch Out on a Practical Platform for Institutional Research Toward Sharing of Its Technology and Knowledge

Shotaro Imai ^{*}, Yoshikazu Asada [†], Akira Itoh [‡],
Toshiki Katanosaka [§], Aoi Kishida [¶], Naruhiko Shiratori ^{||},
Kunihiko Takamatsu ^{*}, Sayaka Matsumoto ^{*}, Masao Mori ^{*}

Abstract

Terenzini classified the intelligence required for an Institutional Research (IR) conductor into three tiers. “Technical and analytical intelligence” is a general technique of statistical analysis but the other two are difficult to learn since they require experience in IR and a deep understanding of the institute to which each belongs. In addition, there is no definitive definition of IR in Japan. These facts confuse IR beginners. To overcome the problem, we launched a platform for sharing technology and knowledge of IR. An IR beginner can access to learn the methodology of IR and an IR expert can share ingenious ideas and techniques through the platform. We named the platform PAIR (Platform of the Art of Institutional Research) and provided it by GitHub. In this paper, we describe the background, the concept, and the future visions of the PAIR.

Keywords: IR human resource development, open data, platform, sustainability

1 Introduction

Institutional Research (IR) has been established and developed in the U.S. since the 1960s. Currently, the definition of IR widely accepted is “Institutional research is research conducted within an institution of higher education to provide information which supports institutional planning, policy formation and decision making” stated in [1], although this statement is general and abstract. The purpose, missions, methods and approaches are not clear. The concrete activities of IR depend on the context of each university.

In Japan, the introduction of Institutional Research (IR) was triggered by the standards of the university establishment in 1991, in which an university assessment system such as

^{*} Tokyo Institute of Technology, Tokyo, Japan

[†] Jichi Medical University, Tochigi, Japan

[‡] Kobe Tokiwa University, Hyogo, Japan

[§] Toyo University, Tokyo, Japan

[¶] Kobe City Nishi-Kobe Medical Center, Hyogo, Japan

^{||} Kaetsu University, Tokyo, Japan

IR is required [2]. While there is no definitive definition of IR in Japan, some explanations of IR are shown in glossaries created by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). In the report [3], IR is mentioned under the term explanation of University Portrait (tentative name) as “By organizing the University Portrait (tentative name), 1) Universities utilize educational information to understand and analyze their activity status, leading to reform (the so-called improvement of IR (Institutional Research) function), . . .” In the glossary of the report [4], the section on “IR (Institutional Research)” is created and defines IR as “Abbreviation for Institutional Research. It refers to a function or department in higher education institutions that conducts investigations and analyses of information related to the institution. By centrally collecting and analyzing institutional information, it enables the institution to smoothly carry out planning, policy formulation, and decision-making. Additionally, it provides institutional information to internal and external parties as needed.” In the glossary of the guideline [5], the section on “Institutional Research (IR) and Educational IR” is created. The definition of IR is the same as the above and the sentence “This guideline specifically addresses the aspects related to education, treating them under the designation of “Educational IR”.” is added to stress on the application of IR to education management.

Generally, IR in Japan is classified into three categories: Educational IR, Research IR, and Management IR, addressing education, research, and management respectively [6]. Educational IR seems to be the most advanced field [6] due to the historical context of the introduction of IR in Japan as we mentioned. Indeed, the establishment of an IR organization is asked in the questionnaire of Comprehensive Support for Private University Reform Project (hereafter Comprehensive Reform) [10]. In Comprehensive Reform, three items are inquired related to IR: (1) Development and enhancement of IR function, (2) Activity content of Educational IR, and (3) IR collaboration with other universities. Especially, there are 12 questions regarding (2), including understanding study time and behaviors, understanding study outcomes, utilization of class evaluations, formulation of assessment policies and evaluations based on them, verification of educational curriculum using IR information, visualization of study outcomes through diploma supplements, operation of verification and improvement cycle of educational curriculum utilizing IR information, admission tracking survey, graduation survey, post-graduation survey, consultation with the industrial sector on study outcome information, and exchange of opinions with corporations on study outcome information (pp. 35-40 in [7]). In addition, there are similar concepts and activities with Educational IR such as Enrollment Management, Institutional Effectiveness [7], and Learning Analytics [9]. Enrollment Management is devoted to the first chapter of the IR textbook [8].

Under the current situation, some of the authors and collaborators started a research project to reveal the academic disciplines and required skills and knowledge of IR from 2022 (JSPS KAKENHI Grant Number 22H00077). The key question of the project is ‘What is the academic significance of establishing IR as a distinct field, and what activities are indispensable to achieve this?’ While there are several reports to introduce international studies about IR skills, investigation of it in Japan and development of educational programs of it are not advanced yet. In addition, there is a critical discussion on the academic significance of IR since IR is a practical methodology of university management, which is administrative work and not academic activity. To respond to the situation, we aim to contribute to the enhancement of IR organizations and cultivating IR conductors by elucidating the academic domain and prerequisite competencies for IR.

To address the challenges, we have focused on the three questions below:

1. Describing a systematic ontology of IR.
2. Identification of a comprehensive IR skill set.
3. Establishing the information infrastructure and technology for IR.

Preliminary findings from our research have already been reported. Matsuda et al. [11] classify English research articles in IR into seven major and 47 minor categories for establishing systemized IR discipline. Iseri et al. [12] report the situation of training-course of IR using the archive data of asagao-ML (which ended and migrated to tulip-ML in Sep. 2022) and found the trends that increase of training programs after 2014, transitions of host organization of training, and a low of practical hands-on workshops. The reports [13, 14] report the results of the questionnaire targeted at Japanese university staff and faculty who work on IR. 47.2 % of the respondents answered their academic background is undergrad. This suggests they are not faculty and researchers as administrative staff. In addition, the specialty of the respondents covers all academic disciplines including not only education, sociology, politics, economics, and information science, which are considered as a deep relation with IR, but also mathematics, natural science, engineering and astronomy. This fact suggests many persons working on IR are not trained and specialized for IR.

2 In Pursuit of Training IR Skills and Knowledge

As we discussed in the previous section, IR in Japan has a wide variety of definitions and interpretations. While this situation shows that IR is active in Japanese universities and covers wide disciplines and academic aspects, it makes IR beginners and IR staff unsure of what to study first. Terenzini classified intelligence to conduct IR into three tiers 1) technical and analytical intelligence, which requires familiarity with the basic analytical processes of institutional research, 2) issues intelligence, which requires knowledge of substantive institutional management issues, 3) contextual intelligence, which requires understanding of the history and culture of higher education in general and of the institute on which one works [15]. Knowledge and techniques for data analysis can be learned from commonly available literature that explains how to use statistical tools and data analysis methods. However, when applying these general statistical skills to a problem occurring in a higher education institution, an IR beginner may face difficulty in selecting appropriate methods and tools. The actual data and its analysis methods cannot be disclosed since the data treated in IR includes sensitive data such as personal and confidential institutional information. For this reason, IR research papers usually discuss abstracted and generalized methodologies, but it is difficult for IR staff with little research experience to apply them to their own work.

In the field of Learning Analytics, J. Kuzilek et al. [16] publish and share anonymized data and present examples of analysis using this. Following this idea, we decided to create a knowledge platform for IR beginners where they can see data close to the data actually used in IR and learn how to use it. Until now, lectures for beginners such as those reported in [12] sometimes include analysis methods using dummy data prepared by instructors, but the datasets analysis methods are not published and shared. Although a few datasets such as [17] have been published, their existence is not well known and they are not utilized. Such datasets and analysis methods are valuable intellectual resources based on the instructor's experience. Based on the ideas of open data and open science [18, 19], we provide our platform as a place to publish and share these datasets and analysis methods as much as possible and aim to disseminate and improve IR technology and knowledge. We named

this platform “PAIR” (Platform for the Art of Institutional Research). Following [20], we present concepts and visions of PAIR and report future expectations and problems in this paper.

3 Platform for IR Technology and Knowledge

In this section, we explain the concepts of the IR technology and knowledge platform “PAIR” and what is currently being realized. The essential concepts can be summarized into two points. One is to provide a first guidance of IR not only data analysis methods (technical and analytical intelligence) but also accompanied by a practical issue and context of a research question (issues intelligence and contextual intelligence) for IR beginners. Here, we assume an IR beginner is a person who has been worked as a university staff but not trained and skilled in IR. The second is to provide a place to share ingenious ideas and techniques for mature IR conductors, who have experience in IR and are in a position to improve the IR community. The detailed concepts will be described in the following subsections.

3.1 Concept of The Platform

To be a guide for beginners of IR, PAIR provides the following:

1. Typical sample datasets of IR,
2. Guidance for data analysis and visualization,

which are starting contents and we will provide any other resources according to the development of the platform. Guidance will include some tips like below to learn practical skills and knowledge:

1. Information of typical data sets on IR fields, such as category and structure.
2. Required skills for data analysis, how to and what to use tools for data analysis.

Especially, choosing analysis tools is a common problem that makes beginners in confusion. To respond to this problem, we will present a process of analysis using several tools for the one sample dataset. Providing explanation of manual for data analysis using programs like R and Python, Business Intelligence (BI) tools such as Power BI and Tableau, and Microsoft Excel. A beginner can learn the skill of analysis with familiar software and computer environment.

We expect mature IR staff to share their ingenious ideas and techniques of IR based on their experience; write explanations for the usage of each tool with their ideas, improve available data, provide original sample data, and suggest ideas to improve the quality of the platform. Whether you are a beginner or not, sharing the experiences of other IR professionals will stimulate mutual learning. The development of the platform will be promoted by encouraging individuals to present their achievements in activities such as creating teaching materials and modifying data using the platform, and by feeding these results back to the platform. Furthermore, we plan to create a link collection of a published dataset, guidance and available websites related to IR, which is another guidance for beginners to know learning resources.

☰ README.md ✎

IR Data Project 🔗

This project aims to enhance the IR (Institutional Research) skills of IR professionals by providing sample data that can be practically used in IR. The sample data distributed is synthetic data generated by tools such as ChatGPT. Moreover, you can contribute to the project by describing data and program code in various languages.

Project Structure 🔗

This project contains several files. Each file contains the content described below:

[README.md](#): Description of the project's purpose, content, and usage.

[datasets](#): Collection of datasets

- `{000-099}.csv` : Datasets for beginners
- `{100-199}.csv` : Datasets for intermediates
- `{200-299}.csv` : Datasets for advanced users

[metadata](#): Information underlying the data and metadata related to data creation

- [creation_process](#): Explanation of the data generation process and technologies used
- [contribution.md](#): Description of how to contribute, including data upload methods and format requirements
- [sources.md](#): Reference information and sources related to data generation and programs

[scripts](#): Programs

- [Excel](#)
- [Python](#)
- [R](#)
- [Tableau](#)


[guidelines](#): Guidelines

- [beginner.md](#): Guidelines for beginners
- [intermediate.md](#): Guidelines for intermediates
- [advanced.md](#): Guidelines for advanced users

License 🔗

- [LICENSE](#): License information for the source code and data.

Copyright (c) Platform for the Art of Institutional Research 2023



All the content including sample data, documents, pictures and sample codes within this repository is licensed under a [Creative Commons Attribution 4.0 International License](#).

Figure 2: Screen shot of the top page of “PAIR” repository. This shows the explanation of the project and the contents of the repository. This explanation is the current situation, and it will be changed according to the improvement of the project.

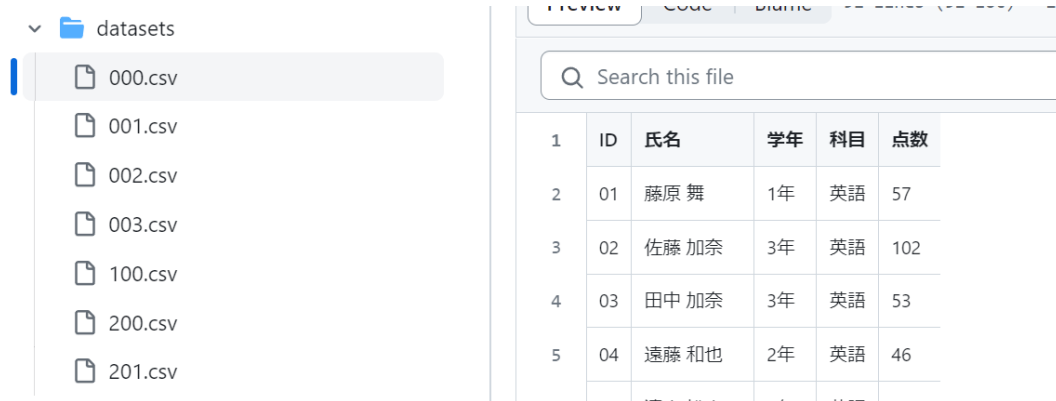


Figure 3: Sample data on PAIR. 000.csv is the simplest (fictional) data including ID, name, year and score of English. This data can be used as a first step of data analysis such as the distribution of the score.

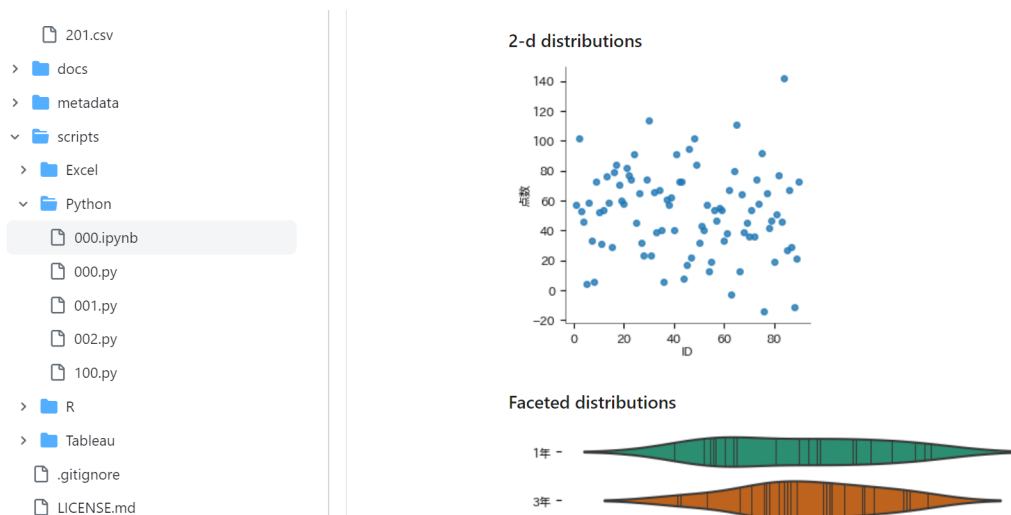


Figure 4: Example of data analysis. We provided a Jupyter notebook using dataset 000.csv as an example.

generative AI, the prompt is a good example and reference for creating synthetic data. The data includes the fictitious student's name, gender, department, grade, score for the course and so on to use students' performance evaluation. Analysis of students' scores is discussed for a beginner level of data analysis for IR in many cases. We provide three types of data to distinguish the level of analytical skill as beginners, intermediates and advanced. This difference in the level is the amount of data header. The data with rich data headers allows rich data analysis such as multivariate analysis.

As an example of analysis, we publish a visualization of the distribution of grade scores using the sample data. Visualization examples and explanations using Excel, Python, Exploratory, R, and Tableau are published (see [21] for details). We provide a Jupyter notebook *000.ipynb* using dataset *000.csv* (see Figure 4). This notebook explains some of the visualizations of the distribution of scores. If you have the environment of Python and Jupyter, the process of data procedure is easily reproduced. We also provided the visualization result process in Microsoft Excel and other tools. Thus, as we discussed previously, readers can follow and reproduce the data analysis using their preferred tool.

Currently, all the contents of PAIR are in GitHub, but we do not limit the resources related to the PAIR to GitHub. For example, a sample dataset of IR or guidance or explanation of IR techniques and knowledge can be published on any other web service. Especially, recent Japanese IT engineers favor Qiita [25], Note [26] or video-sharing services like YouTube [27] to share their know-how. PAIR will provide a collection of web links to such external resources.

We provide all the contents including sample data, explanations, and sample codes are shared under the CC-BY 4.0 license (©Platform for the Art of Institutional Research 2023). We allow to utilization of the contents to be quoted, redistributed, and modified as far as the license allowance. It means you to not only use the sample data and explanations provided by PAIR for learning but also use and refer to them in your lectures and research. You can also modify our resources to add your ideas. If you share these modifications with us, it will be helpful to improve the quality of the platform. To fetch in your modification, you can use Pull Request or contact us directly.

As mentioned above, IR experts may create their sample data and use it in lectures, etc., or make it publicly available. In such cases, we are welcome to collaborate or make links. The best way to make such collaboration has not yet been determined, but we start by linking to a personal site using a URL link, or by importing sample data into the PAIR repository by the marge function of Git. Some of you may want to create a new platform project inspired by PAIR. We also welcome such derivation because our essential concept of PAIR is to activate the IR community and improve the quality of IR works in Japan. This derivation is realized by the fork function in Git. We expect any new activities using the contents provided by PAIR such as modification or adding new sample data, and explanation of data analysis will be published and reported as your experience. This is the right for everyone, not just the authors but others involved in PAIR. We hope that giving everyone the right to publish will revitalize the IR community and improve the quality of IR activities.

4 Challenges Toward Sustainable Operation

This platform is still in the start-up phase, so there are many problems with its sustainable operation. The first one is the establishment of the organization team. The members and

roles are not clear yet. The authors will serve as the initial operations team, managing the GitHub page and working to improve the data and enrich the explanations of IR techniques. However, a small number of members is not suitable for a long-term operation. The operation cannot be sustained with volunteers, it is necessary to organize the management members. It is necessary to learn know-how for the operation of an open platform from references such as [28].

The quality of sample data is another problem. GitHub is used often for open-source software development projects, and it is good at managing a team development of program source code. We aim to ensure quality by establishing rules such as rules for data merging and standardization of Commit message formats (e.g. [29]).

The most important is to enrich the contents of the website. Improvement of sample data and explanations is the first task. Furthermore, we need to collect and increase the link connections with other websites and services related to IR. As we mentioned above, GitHub is good at the management of source code but not at documentation and connection with other services. Therefore, we may need to move the main service platform to another knowledge management service. For example, Notion is a recent trend of knowledge base [30]. It has functions of wiki-like documentation with a markdown format and simple database and allows integration with other services including GitHub. It is possible that we create a public page on Notion and it can be a home page of PAIR. The explanations of data analysis and any knowledge of IR and link collections are managed by Notion and the sample dataset is managed by GitHub. We will redesign the structure of our page and the way of providing the content in the near future.

5 Summary

In this paper, we declare to launch a platform to acquire a practical technique and knowledge of Institutional Research and it is named “Platform of the Art of Institutional Research (PAIR).” The aim and concept of PAIR are presented. To understand the demand for this platform, we discussed the situation of IR in Japan from the perspective of the diversity of the activities and the diversity of IR staff. We found that the missions and activities of IR are related to higher education policy, and it is difficult to define it unambiguously. The IR staff have a diverse range of work experience. There is a case in which a person who has no experience in academic research is in charge of IR, and the person gets confused and isolated. To respond to such a situation, we decided to create the platform to be a first guidance.

The platform publishes knowledge about IR with specific sample data and explanations of analysis methods for each tool. This platform will become a place where anyone can freely use the provided content and add any new ideas and resources. Using this image as a sandbox, we explained that each person can freely learn and contribute by creating various statues using the sand, explaining how to make them, and adding new sand.

Although we still have problems of sustainability such as how to operate the platform and how to maintain and manage the quality of content, our proposal and creation of such a platform is an important contribution to the promotion of IR. We hope PAIR will be a place of learning and teaching, a place to get experiences in IR activities and an entrance to the IR community.

Acknowledgments

This work is supported by JSPS KAKENHI Grant Number 22H00077. The authors appreciate Tsunenori Inakura (Tokyo Institute of Technology) and other members of the JSPS KAKENHI for supporting the launching of the platform and generating sample data. We also thank the members of the Japan Association for Institutional Research for fruitful discussions on the management of the platform.

References

- [1] J. L. Saupe, “The functions of institutional research. 2nd Edition,.” 1990.
- [2] M. Fujiwara, S. Chikamori, A. Asano, and N. Soshii, “Development of institutional research (IR) to support policy formulation in the educational field: Definition, organizational structure, and processes of institutional research in the educational field at Ritsumeikan University (Japanese),” *Univ. Adm. Stud.*, vol. 4, pp. 17–31, 2009, doi:10.34382/00005804.
- [3] MEXT, “Transforming the Quality of University Education for Building a New Future: Towards Universities that Cultivate Lifelong Learning and Autonomous Thinking—Transforming the Quality of University Education for Building a New Future: Towards Universities that Cultivate Lifelong Learning and Autonomous Thinking (Japanese),” 2012. [Online]. Available: https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325047.htm. [Accessed: 01-Jun-2023].
- [4] MEXT, “Grand Design for Higher Education Towards 2040 (Japanese),” 2018. [Online] Available: https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1411360.htm. [Accessed: 15-Sep-2022].
- [5] MEXT, “Educational Management Guidelines (Japanese),” 2020. [Online]. Available: https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1411360_00001.html. [Accessed: 01-Jun-2023].
- [6] R. Yamada, “Trend of IR in Japan: Emergence of IR Including Management, Teaching and Learning and Research,” *Proc. Inst. Stat. Math.*, vol. 68, no. 2, pp. 197–208, 2020.
- [7] Y. Aioi, S. Iseri, S. Imai, T. Oishi, S. Okada, N. Kondo, T. Sugihara, S. Tajiri, K. Tsukamoto, Y. Tsubakimoto, K. Nishiyama, T. Matsuda, and M. Mori, “University IR Standard Guidebook: Know-how and Practice of Institutional Research - Paperback.” Impress R&D, 2022.
- [8] R. D. Howard, “Institutional Research: Decision Support in Higher Education,” Association for Institutional Research, 2001.
- [9] T. Matsuda and Y. Watanabe, “Pedagogical IR, Learning Analytics, and Educational Technology,” *Information and Technology in Education and Learning*, vol. 41, no. 3, pp. 199–208, 2017, doi:10.15077/jjet.42028.
- [10] MEXT, “Comprehensive Support for Private University Reform Project,” 2020. [Online]. Available: https://www.mext.go.jp/a_menu/koutou/shinkou/07021403/002/002/1340519.htm. [Accessed: 01-Oct-2023].

- [11] T. Matsuda, T. Sugihara, and M. Ishii, “Academic Fields Covered by IR — From the Results of the English Literature Search —,” in 11th Meeting Japanese Institutional Research, 2022, pp. 170–173, doi:10.50956/mjir.11.0 170 33.
- [12] S. Iseri, T. Oishi, N. Kondo, and K. Takamatsu, “Actual Conditions of Training Opportunities on IR in Japan — From the Archived Data of Asagao ML —,” in 11th Meeting Japanese Institutional Research, 2022, pp. 164–169, doi:10.50956/mjir.11.0 164 32.
- [13] S. Iseri, T. Oishi, N. Kondo, and K. Takamatsu, “Training Opportunities and Required Skills for Institutional Research Programs in Japan,” 23rd Annual SEAAIR Conference Proceedings, 2023, in press
- [14] M. Mori and S. Iseri, “Survey Research for Building the Academic Foundation and Human Resource Development in University Institutional Research,” in 26th Japanese Association of Higher Education Research, 2023, pp. 141–144.
- [15] P. T. Terenzini, “On the nature of institutional research and the knowledge and skills it requires,” *Res. High. Educ.*, vol. 34, no. 1, pp. 1–10, 1993.
- [16] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open University Learning Analytics dataset,” *Sci. Data*, vol. 4, no. 1, p. 170171, 2017 doi:10.1038/sdata.2017.171.
- [17] M. Mori, “On the Sample Data Set for Analysis of Educational Affairs,” in 12th Meeting Japanese Institutional Research, 2023, doi:10.50956/mjir.12.0 78 1.
- [18] N. Miyairi, “Open science and scientific data,” *J. Inf. Process. Manag.*, vol. 57, no. 2, pp. 80–89, 2014, doi:10.1241/johokanri.57.80.
- [19] H. Koshiba and K. Hayashi, “Open Data for Open Science,” *Journal Artif. Intell.*, vol. 31, no. 2, pp. 262–268, 2016.
- [20] S. Imai, Y. Asada, A. Ito, T. Katanosaka, N. Shiratori, K. Takamatsu, S. Matsumoto and M. Mori, “Start-up a Practical Platform for Institutional Research,” in 12th Meeting Japanese Institutional Research, 2023, doi:10.50956/mjir.12.0 66 1.
- [21] IR-Platform, “Platform of the Art of Institutional Research,” 2023. [Online]. Available: <https://github.com/IR-Platform/PAIR/>. [Accessed: 01-Oct-2023].
- [22] DAIR.AI, “Prompt-Engineering-Guide,” 2023. [Online]. Available: <https://github.com/dair-ai/Prompt-Engineering-Guide/>. [Accessed: 01-Oct-2023].
- [23] kaityo256 (H. Watanabe), “sevendayshpc (Become a Supercomputer Programmer in a Week!),” 2019. [Online]. Available: <https://github.com/kaityo256/sevendayshpc/>. [Accessed: 01-Oct-2023].
- [24] OpenAI, “ChatGPT: Chat Generative Pre-trained Transformer.” [Online]. Available: <https://chat.openai.com/>. [Accessed: 01-Oct-2023].
- [25] Qiita, “Qiita (A service for recording and sharing knowledge related to engineering).” [Online]. Available: <https://qiita.com/>. [Accessed: 01-Oct-2023].
- [26] Note, “note (Media platform),” 2023. [Online]. Available: <https://note.com/>. [Accessed: 01-Oct-2023].

- [27] YouTube, [Online]. Available: <https://www.youtube.com>, [Accessed: 01-Oct-2023].
- [28] GitHub, “Open Source Guides (Starting an Open Source Project),” 2014. [Online]. Available: <https://opensource.guide/ja/starting-a-project/>. [Accessed: 1-Oct-2023].
- [29] joshbучea (J. Buchea), “semantic-commit-messages.md,” 2017. [Online]. Available: <https://gist.github.com/joshbучea/6f47e86d2510bce28f8e7f42ae84c716>. [Accessed: 01-Oct-2023].
- [30] Notion Labs, Inc., “Notion,” [Online]. Available: <https://www.notion.so>. [Accessed: 20-Aug-2023].