On-Campus Data Utilization: Working on Institutional Research in Universities

● Aiko Otomo ● Yutaka Iwayama ● Takao Mohri

Institutional research (IR) in universities refers to a wide range of activities such as collection and analysis of on-campus data, planning of measures for improvements, and implementations and verifications of the measures for the purpose of management improvement, student support, and higher-quality education. In Japan's universities, improvements with on-campus data utilization have been increasingly expected and there has been a big upsurge in IR for this. And universities are seeking approaches for specific practices. In response to such trends in Japan, Fujitsu is currently working on IR research and study, aiming to provide new services that can contribute to university business and management. Specifically, we carried out surveys on the needs for and expectations of IR, made hypotheses, and conducted trial experiments of data analysis. By using various statistical analysis methods, we also tested the trend analysis of the improvement in education quality and dropout prevention, for which data utilization will be especially needed in universities. Furthermore, research on dropout prediction was carried out, by which it is possible to predict which students may leave universities in the future from their current situation. Through the approaches above, we are currently defining the concept of Fujitsu's IR and examining the provision of services supporting IR practices in universities. In this paper, the IR trend research and study that we have tackled so far and Fujitsu's IR are described.

1. Introduction

Institutional Research (IR) in universities refers to a wide range of activities such as collection and analysis of on-campus data, planning of measures for improvements and implementations and verifications of the measures for the purpose of management improvement, student support and higher-quality education.[1] It has long been in place in foreign countries including the U.S. and plays an important role as a function to support analysis and decision-making.

In Japan, an understanding of the necessity and effectiveness of IR is becoming widespread, as exemplified by advanced activities including the "Four-University IR Network"[2] led by Doshisha University and "Enrollment Management IR"[3] of Yamagata University.

In the "Development of University Portrait" (tentative name)[4] of the Subdivision on Universities, Central Council for Education, Ministry of Education, Culture, Sports, Science and Technology, infrastructure is being developed to disseminate information and assist in the utilization and publication of educational information of universities. There are high expectations for on-campus data utilization in Japan, and momentum is increasing for this.

Fujitsu has long supported operations in universities in terms of systems including administration, library and finance ones. Based on the know-how of operations developed up to now, we are currently working on IR research and study, aiming to provide new services that can contribute to university business and management.

This paper describes the IR trend research and study in which we have been engaged up to now and Fujitsu's IR.

2. Surveys on needs for IR

As part of the research and study, we first conducted interview surveys on the needs and

expectations for IR and issues with its introduction. Survey respondents were the media center, information center, planning department, management strategy department, etc. of a total of 22 universities. There are two major points that are focused on in the interview:

- What can be done by data utilization
- Issues with working on data utilization

2.1 What can be done by data utilization

The results of the interview have shown that improvements in education quality, student support/career development and entrance examination strategy are attracting major interest. The following shows some of the needs under the respective items.

1) Improvement in education quality
- Relationship between the order of course-taking and maximization of students' performance
- Instructor evaluation, class evaluation and curriculum evaluation
- Effect of extra-lecture programs (such as studying abroad) on students

2) Student support/career development
- Understanding of the process that leads to dropping out and dropout prevention
- Understanding of the relationship between career awareness formation and performance/course-taking
- Presentation of comparison with students of other universities and improvement of students' motivation

3) Entrance examination strategy
- Public relations (PR) activities aimed at high schools, securing of excellent students and publicity to the provinces
- Effect on campus visit participants' taking of the examination

2.2 Issues with working on data utilization

As a result of the interview, it has been shown that respondents find issues with information infrastructure, organizational culture, specialized ability and specialized organization. The following shows (some of) the issues under the respective items.

1) Information infrastructure
- Framework allowing centralized management of data
- Framework allowing provision of necessary data for those who need them

2) Organizational culture
- Difficulty in coordinating opinions between departments, instructors, etc.
- Interorganizational barriers for data collection and all-campus analysis
- Guidance for teachers who have not yet used the system to show them how to use it

3) Professional expertise
- Specialized skills including analysis ability and database management

4) Specialized organization
- Difficulty in launching a specialized organization that practices IR
- Limits of workload of the information system department

The results of the interviews show that there are increasing expectations for university improvement by on-campus data utilization. However, in practicing data analysis, universities are aware of the issues with and limit to information infrastructure development, data collection and the need to have people with a professional expertise. Thus, solutions to address these issues are required.

3. Trial of data analysis

In order to verify what types of analysis and visualization are possible in reality, we performed a trial experiment of data analysis. There are two points to work on in the trial experiment:

- Hypotheses extraction
- Trial experiment

3.1 Hypotheses extraction

For items that proved to attract especially high levels of interest in the results of the interview, we studied the issues that universities are faced with and extracted hypotheses for them. We conducted brainstorming sessions and questionnaires among Fujitsu personnel engaged in education such as sales staff and systems engineers, and extracted about 400 source materials of hypotheses. We then eliminated redundant and unanalyzable ones to organize them into 100 hypotheses. Those hypotheses were evaluated by two universities and were discussed in terms of their adequacy, priority, importance, etc.

3.2 Trial experiment

We used sample data to analyze and visualize the hypotheses created and studied the methods of verifying them, analysis procedures, person-hours required and analysis tools. **Figure 1** shows the results of a trial.

In an attempt to actually analyze data, we have found out that hypotheses as they are do not allow visualization in graphs. For visualization, it is necessary to verify the purpose of a graph preparation based on a hypothesis and discuss the axes and representation of the graph. For example, we considered a hypothesis that "students admitted by Admission Office (AO) entrance examinations or recommended by designated schools have lower academic ability and those admitted by general entrance examinations have higher academic ability." Here, the axes of analysis must be defined before the hypothesis can be visualized: Students need to be classified by the form of admission to find a grade point average (GPA) for each semester and classification to plot a line graph with the semester and GPA as the horizontal and vertical axes respectively.

Actual visualization of data may show results that are in line with the hypothesis, that are different from the hypothesis or that are unexpected. Hypotheses are subjective views or speculations based on experience and visualization of facts by using data may back up, or conversely discredit the hypotheses. In universities, hopes and momentum have been rising by measures based on analysis of data.

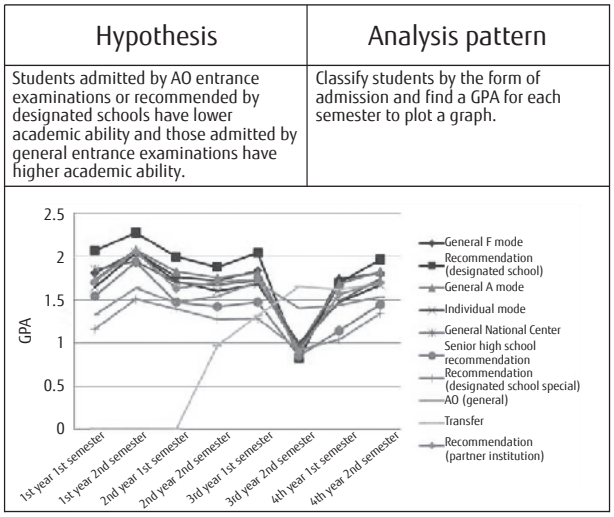

Figure 1
Result of trial analyses.

4. Dropout prevention by IR

Among the needs listed in the "Surveys on needs for IR" section, there was an especially strong need for "dropout prevention" under "2) Student support/career development" that universities intended to work on as soon as possible as the most pressing issue. Accordingly, we used various statistical analysis methods to try analyzing the trends in relation to dropout prevention and conducted research on "dropout prediction" to predict which students may leave universities in the future based on their current situations. The following outlines the research.

The overall flow of the approach to preventing dropout by utilizing data is shown in **Figure 2**. Data from various systems that exist on campus are analyzed to identify students requiring attention or about to drop out, out of the enrolled students. Data that can be used include information before admission to universities, entrance examination information, post-admission enrollment information and learning activity information such as information about courses taken and performance, attendance information and reports, many of which are expected to be useful in analysis. After students requiring attention or about to drop out have been identified, individual interviews, counseling and support are provided for those students to prevent them from dropping out. The results of counseling, etc. can be stored in a student record system (Campusmate-J/StudentChart), etc. to use for continuous follow-up activities.

5. Dropout prediction method

In order to identify students requiring attention or about to drop out, out of the enrolled students, it is necessary to predict the risk of dropping out for each student. On the assumption that students of the same university (faculty or department) show a similar trend concerning the risk of dropping out even in different years, we studied a method of predicting the possibility that the present students will drop out or not by using data on students in the past.

To develop the dropout prediction method, we first visualized dropout-related data and gained an understanding of the trends for discussing the prediction method. The following presents an example of visualization of dropout-related data of a certain university. **Figure 3** shows the dropout rates (percentage of the

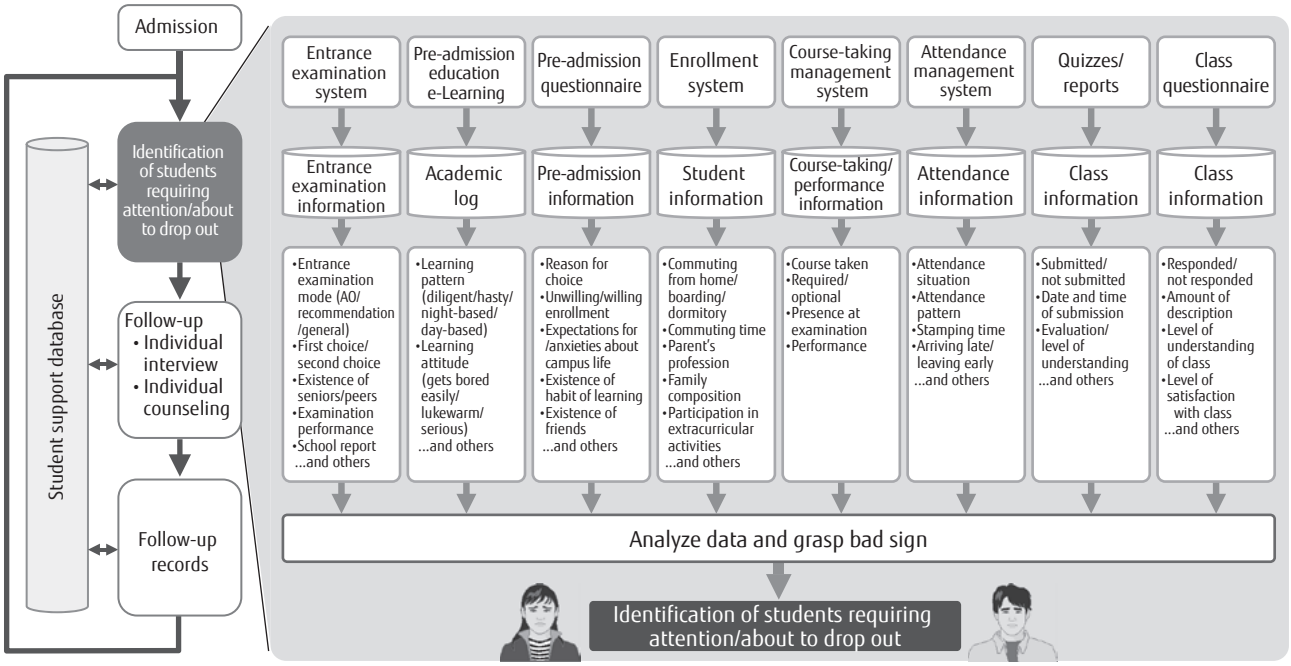

Figure 2
Overall structure of approach to dropout prevention by IR.

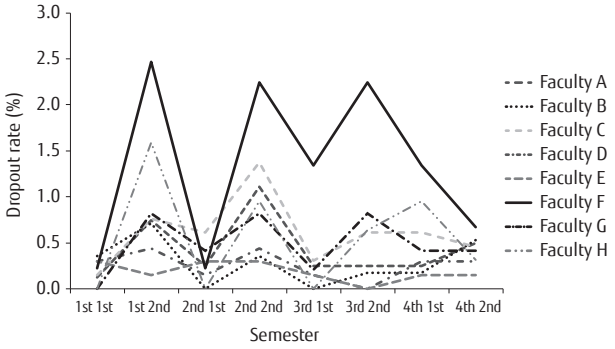

Figure 3
Dropout trend for each faculty/semester.

number of students who dropped out in the semester out of the number of students enrolled in the faculty in the same year) for the individual faculties and semesters. This indicates that the dropout rates may differ for individual faculties even within one university. Within the same faculty, the dropout rates depend on the semester and the rate has proved to be high in the second semester of the first year and the second semester of the second year.

Figure 4 shows the trend of dropouts for each entrance examination mode. It reveals that, in this university, the dropout rates of the students who passed the general entrance examination are high.

Figure 5 shows the relationship (correlation coefficient) between the performance of each course in the first year and whether the students who took the course dropped out in the first year. The more negative a correlation coefficient that a course has (such as the courses surrounded by a dotted line in the figure), the more likely it is that students will drop out from that course when the obtained score is low. The vertical axis shows the number of courses that have the correlation coefficient shown on the horizontal axis. Courses that give a higher probability of dropping out include both required and optional courses. Optional courses that may have an influence on dropping out in spite of the fact that they are not required can be assumed to be significant courses.

6. Development of dropout prediction method and example of prediction

We have developed a method of predicting students who are likely to drop out based on the results of these visualizations. For prediction, this method takes into consideration the difference in the trend of dropping out depending on the university, faculty,

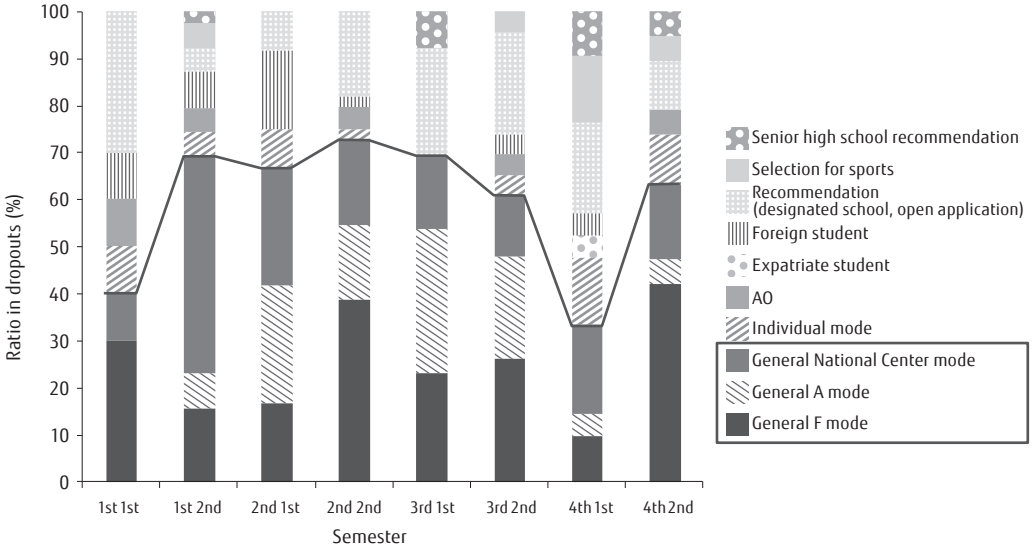

Figure 4
Ratio in dropouts for each semester/entrance examination mode.

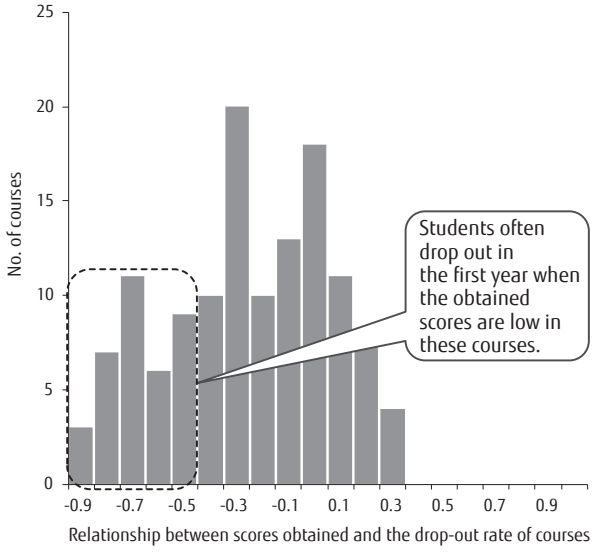

Figure 5
Relationship between correlation coefficient for each course and number of courses.

department and semester. It particularly picks up courses that may have an influence on dropping out to use for prediction. As a method of prediction, we have used machine learning, specifically a support vector machine (SVM). As data to use for prediction, we have chosen information at the time of admission (average evaluation in high school and entrance examination type [general, AO, etc.]) and information about courses taken and performance from the viewpoint of the ease of data availability.

Table 1
Example of prediction results.

Prediction	Recall ratio (%)	Precision ratio (%)
Drop out by end of second year?	35	95
Drop out by end of fourth year?	26	85
Repeat a year?	23	37

As an assumed scene of prediction, we have used the time of completion of registration for the second semester of the first year and predicted whether the student will drop out based on the data available up to that point. The prediction was made from three perspectives: drop out by the end of the second year, drop out by the end of the fourth year or repeat a year.

The predicted results are shown in **Table 1**. The prediction of dropping out in the second or fourth year shows a high precision ratio, while the recall ratio is low. This indicates that students predicted to drop out actually drop out with high probability (high precision ratio) but the prediction failed to include many of those who actually dropped out (low recall ratio). For repetition of a year, both recall and precision ratios are low. We think that attributes that characterize repetition of a year have not been sufficiently detected (used) yet, which is a challenge to be tackled in the future.

At present, we are evaluating the prediction method and brushing up the technique in ways such as adding attribute values of students. To further improve the accuracy, we intend to expand data used for prediction to utilization of students' behavior in classes (such as attendance and submission of reports), pre-admission information and questionnaire information as shown in Figure 2.

7. Fujitsu's IR

We have summarized the concept of IR provided by Fujitsu through surveys on needs as well as trial and study of analysis (**Figure 6**). Fujitsu has provided various systems that have supported university operations up to now and, for IR, data from those systems can be collected for analysis and visualization according to customers' issues. We are considering ways to provide an analysis platform for that purpose and to offer services of analysis itself.

We have also worked out the IR implementation steps through trial experiments of data analysis (**Figure 7**). Possible steps of IR implementation include: current state/status analysis/identification of issues, drafting of hypotheses, examination of analysis methods, construction of the analysis platform, data collection, data analysis, verification of analysis results and drafting of improvement measures. We think, however, that these steps do not necessarily have to be taken in this order and universities can start with what is possible. We believe that it is important to continuously make efforts while gradually expanding the scope of implementation: make a small start to obtain output and build a consensus within the campus in the meantime and attempt to solve the issues while working on data analysis.

As the data analysis platform, data from various on-campus systems can be aggregated into a database to use for data analysis, visualization of results (report) and provision of the platform for utilization and disclosure (**Figure 8**). The systems to collect the data from can be not only Fujitsu's solution products but also products of other companies or Microsoft Excel or

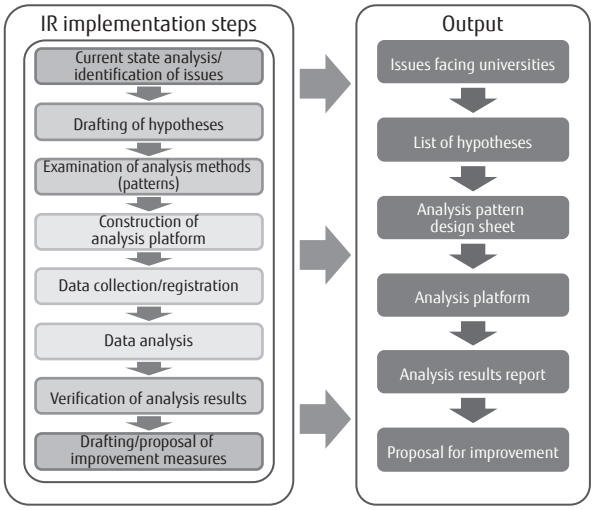

Figure 7
IR implementation steps.

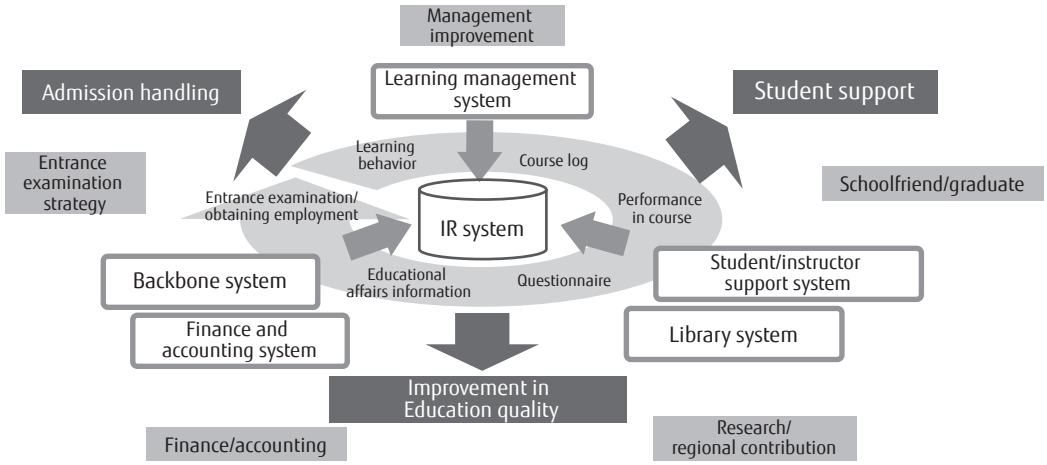

Figure 6
Concept of IR provided by Fujitsu.

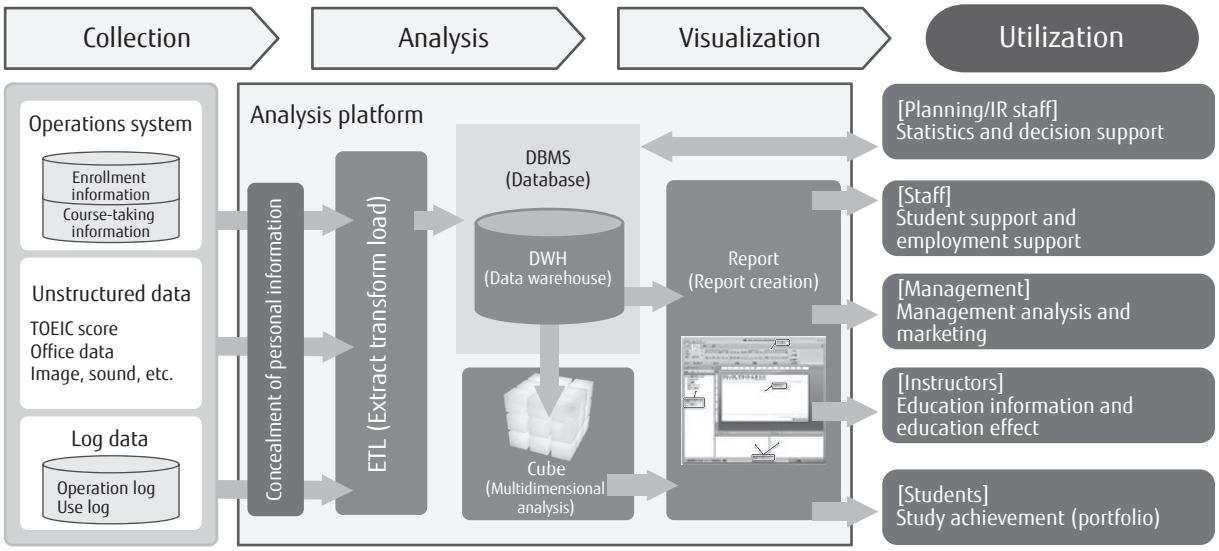

Figure 8
Image of analysis platform.

other types of data. The data can be utilized for the planning department, staff, management, instructors, students, etc. and flexible provision of analysis results according to the respective needs should be necessary. In particular, to expand data analysis to student support in the future, we think that attendance information acquired from the attendance management system (Campusmate-J/Attendance Management) and learning activity information acquired from the learning management system (CoursePower) will become important.

For data analysis and visualization, we support analysis according to the requirements and requests of universities and also propose the analysis methods that we have worked on up to now in research and study. Furthermore, we plan to offer in the near future a solution for dropout prediction that is currently under development.

8. Conclusion

This paper has described trend research and study of IR in universities and Fujitsu's concept of IR. We expect IR will become increasingly necessary and important for universities to explore formulation of strategies for their survival. In the future, we intend to plan, develop, and provide the analysis platform that supports IR implementation by universities and services in relation to analysis support under the concept of Fujitsu's IR.

References

1) K. Oki and S. Okada: Deta ni Yoru Daigaku Kyoiku no Jiko Kaizen (Self-Improvement of University Education by Means of Data—Past, Present and Future Outlook of Institutional Research). Gakubunsha, 2011 (in Japanese).
2) The Universities Institutional Research Consortium (in Japanese).
http://www.irnw.jp/
3) Microsoft: Example of introduction into Yamagata University (in Japanese).
http://www.microsoft.com/ja-jp/casestudies/yamagata-u.aspx
4) Ministry of Education, Culture, Sports, Science and Technology: Development of University Portrait (Tentative Name) (in Japanese).
http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/015/attach/1316454.htm

Aiko Otomo
Fujitsu Ltd.
Ms. Otomo is currently engaged in service planning and development for IR in universities.

Yutaka Iwayama
Fujitsu Laboratories Ltd.
Mr. Iwayama is currently engaged in research and development for data analysis and prediction method for IR in universities.

Takao Mohri
Fujitsu Laboratories Ltd.
Dr. Mohri is currently engaged in research and development of education solutions making use of ICT.