An Architecture of the Information Collected from Higher Education Institutions by NIAD-QE

Kei Kurakawa [*], Syun Tutiya [†]

Abstract

We reconsider the architecture of the information collected from higher education institutions by the National Institution for Academic Degrees and Quality Enhancement of Higher Education (NIAD-QE) and analyze its character. NIAD-QE periodically collects educational information, evidence-based documents, and self-assessment reports to enhance the quality of higher education, which are in a broad sense used to support university reform. These data represent a collection of organizational information gathered by observing higher education institutions in which, to collect the facts about organizational members, the data are aggregated in tabulated form. However, we find that this is not as efficient a way of conveying the information content in a limited data size as the original individual data sets.

Keywords: analytical power, higher education data, information architecture, institutional-level data, personal-level data

1   Introduction

"University reform" for Japanese higher education institutions started in 1991. Since then, regulations for higher education institutions and academic sector agencies have changed continuously [1]. In the first year of the reform, the standards for establishing universities [2] were de-regulated to allow higher education institutions flexibility in providing academic programs; it was also stipulated that they should strive to implement self-assessment as part of quality assurance for higher education. Subsequently, in 1998, self-assessment was mandated, so the universities had to implement it and publicize the results of the assessment reports. A third-party evaluation was also recommended in the standards.

In 2004, Japanese national universities and inter-university research institutions became corporations under the National University Corporation Act [3]; thereafter, they became responsible for the management of the corporate bodies and were required to be evaluated by the National University Corporation Evaluation Committee every five years. In the same year, a certified evaluation and accreditation system was established for higher education institutions under the School Education Act [4]. Higher education institutions were required to be certified at least once every seven years. The National Institution for Academic Degrees and Quality Enhancement of Higher Education (NIAD-QE) was designated as a certified evaluation and accreditation organization for the system and was called on by the National University Corporation Evaluation Committee to discuss and report on the evaluation of each university corporation's

[*] National Institution for Academic Degrees and Quality Enhancement of Higher Education, Tokyo, Japan, ORCID: 0000-0002-7031-1846
[†] National Institution for Academic Degrees and Quality Enhancement of Higher Education, Tokyo, Japan, ORCID: 0000-0003-2790-4846

attainment of its mid-term objectives for education and research.

In 2011, the Enforcement Regulations for the School Education Law [5] were amended to require higher education institutions to disclose educational information publicly. An expert council at the Ministry of Education, Culture, Sports, Science, and Technology, Japan (MEXT), proposed a policy for promoting the utilization and disclosure of academic information in universities. Under the amendment to the law and the new policy, in 2015 NIAD-QE started to provide an information service, namely Japanese College and University Portraits [6], to deliver information about the educational activities of universities and junior colleges.

The two types of evaluation systems for higher education institutions mentioned above, i.e., the certified and accreditation evaluation system and the National University Corporation Evaluation system, as well as the Japanese College and University Portraits, require NIAD-QE to collect, organize, and publicize information from higher education institutions for every year or applicable period.

The collection cycle for the information depends on NIAD-QE's business units, and the coverage of each of the collections of information unfortunately overlaps with or includes that of other collections. Staff on the university side in charge of providing information often complain about the inefficiency of the information collection system and sometimes complain about its low level of utilization.

We now reconsider the architecture of the information collected from higher education institutions by NIAD-QE and analyze its character. In the following sections, we list NIAD-QE's collections of files and databases from higher education institutions, then describe their information architecture with reference to related work, and finally conclude our paper with a mention of future work.

2   File and Database Collections to Enhance the Quality of Higher Education

NIAD-QE periodically collects educational information, evidence-based documents, and self-assessment reports to enhance the quality of higher education, which in a broad sense are used to support university reform. The list below shows such documents and relevant information services in each business category, which can be accessed either by any member of the public or only by a limited set of stakeholders.

- Basic School Survey equivalent data collection (for national universities, public universities, for-profit universities, and public junior colleges)
    - o Basic university information
    - o University information utilization system (a business intelligence tool)
- Certified evaluation and accreditation (for universities, colleges of technology, and professional graduate schools)
    - o Self-assessment reports
    - o Evaluation reports
    - o Evidence (materials and data)
    - o Common basic data formats
- National university corporation evaluation (for national universities and inter-university research institute corporations)
    - o Performance reports
    - o Performance evaluation reports
    - o Evidence (materials and data)

- o Input data collection, data analysis collection
- o Judgment of the level of research achievement
- Japanese College and University Portraits (JPCUP) (for national universities, public universities, for-profit universities, and public junior colleges)
  - o JPCUP for domestic public release
  - o JPCUP for international public release
- Other business categories
  - o Higher education institutions database search system (for universities, junior colleges, colleges of technology, professional training colleges, and educational institutions, as determined by ministries or agencies) by National Information Center for Academic Recognition, Japan
  - o Survey of public universities
  - o National university corporations' financial statements

For each of the above business categories, documents and data are collected in a variety of computer resource formats. Most documents are still stored on cloud storage in their original formats, such as Word documents, Excel files, and PDFs. Some are transformed into PDF files for internal use or public disclosure. Some educational data collections are loaded into a database to provide search functionality and display the contents through a user-friendly interface over the Internet. Recently, part of the database has been connected to business intelligence tools, i.e., Power BI and Tableau, which can be accessed by a limited set of users who have provided their institution's data in order to analyze the facts about their institution relative to all institutions through a visually appealing and analytical web interface.

3 The Information Architecture of the File and Database Collection

All of the above files and data on education and research are collected from higher education institutions in accordance with each business purpose, organized into databases, calculated to derive statistics, and publicized on the Internet as a service with basic web search functions. To make efficient use of the files and data collection backend and to sufficiently harmonize the services to achieve a high level of utilization, it is important to consider the information model as a first-class property in the knowledge domain.

3.1 Organizational Information Collection

NIAD-QE collects all types of information resources at the organizational level that refer to either the organization as a whole or a sub-organization.

As illustrated in Figure 1, operational circumstances and states of organization as inputs, activities, and outputs, as well as matriculation and graduation in the context of higher education and research, must be observed. Activities are examined at the admission, curriculum, and diploma stages from an educational perspective, as are those research processes that are concretely integrated into educational processes.

Each organization refers to a set of persons. As illustrated in Figure 2, an organization may include sub-organizations. The largest organization is the entire higher education institution. A faculty is a sub-organization of the institution that may include several departments. A professor

may belong to more than one organization under a cross-appointment, resulting in an overlap between the two organizations. A diploma program for students is generally conducted under a faculty. This situation means that the faculty corresponds to a group of teachers and students. There are also cases in which a diploma program for students is offered under more than one faculty. This situation means that the diploma program corresponds to a group of students that is different from the group of teachers. Thus, an organization refers to a set of persons whose constituent members vary depending on the management policy of the institution.
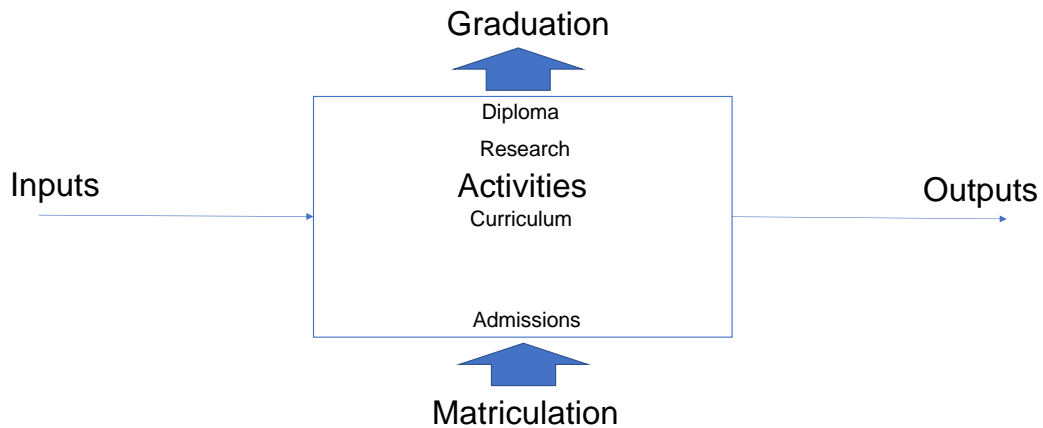

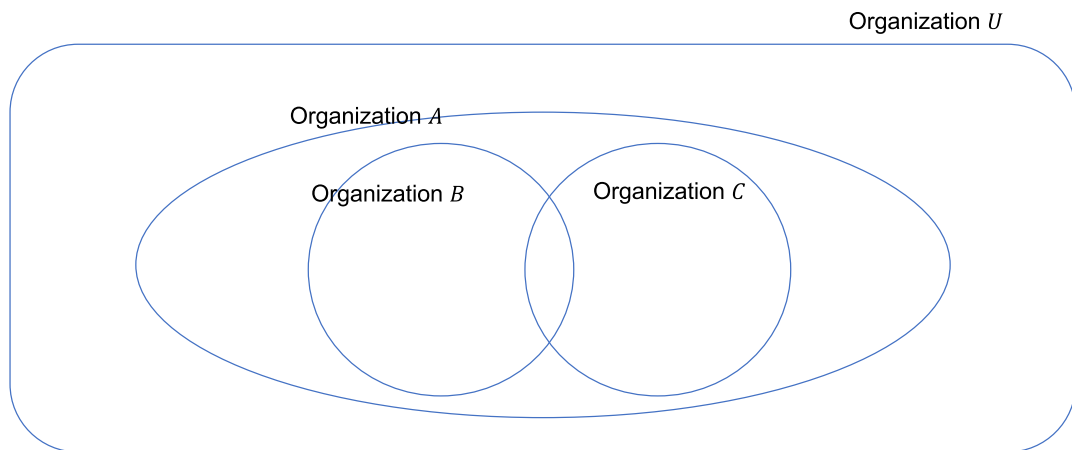Figure 1: Information model for observing higher education institutions

Figure 2: Organizational structure

3.2   Observation Items

NIAD-QE surveys organizations according to the information model shown in Figure 1. It asks the organizations about themselves, then asks for facts from the perspectives of inputs, activities, outputs, matriculation, and graduation, and finally asks for financial statements. Each aspect of the information model can be broken down into the concrete aspects below.
- Organization
   - Institution

- - o Faculty
  - o Education and research programs
- Inputs
  - o Members of the organization
    - Teachers
    - Clerical staff
  - o School sites, facilities, and equipment
    - Empty spaces
    - Athletic grounds
    - School buildings
    - Books and libraries
    - Attached facilities
    - Machines, instruments, and specimens
  - o Research inputs
- Activities
  - o Education policies
    - School mottos
    - Admissions policies
    - Curriculum policies
    - Diploma policies
  - o Education services and support
    - Curricula
    - Extracurricular programs
    - Student financial support
  - o Students' learning status
  - o Research activity status
- Outputs
  - o Research outputs
- Matriculation
  - o Entrance exams
  - o Students
- Graduation
  - o Graduate outcomes
- Financial statements

3.3 Tabulated Data Collected from Higher Education Institutions

When it comes to fact-finding surveys of the members of organizations, i.e., establishing the facts about the teachers, clerical staff, and students in these organizations, NIAD-QE has been accustomed to aggregating the facts in a tabulated data format that characterizes the survey method.

The Basic School Survey, which has been conducted by MEXT every year since 1948, is the model for how all business units in NIAD-QE build data collection systems. In the National University Corporation Evaluation process, the survey sheets that are used to aggregate institutional data as "input data collection" were put together as an extension of the Basic School Survey sheets, technically and contextually.

The institutions should already have a list of organizational member profiles as an original

database for their administrative work. Then, they must calculate the number of members according to the survey requirements, using their original database, and produce tabulated data. Tabulation from an original database requires a process of calculation to count the corresponding elements in the data set according to tabulation properties. This is a demanding task for the staff in charge of the work.

In recent years, the number of sub-organizations in higher education institutions has tended to increase because "university reforms," along with social demands and changing environments, have stimulated institutions to set up academic programs. Teachers' academic role types vary according to the flexibility allowed for establishing academic positions. Student types also vary, as institutions have been able to provide students with flexible courses to attain academic degrees. These changing circumstances have resulted in an increase in the complexity of personal data attributes and the number of tabulated data sheets the institutions need to return.

3.4 Analytical Power of the Original Personal Data Compared to the Tabulated Data

To understand the survey method from an analytical power perspective, we conduct an analysis of its representational capability compared to the alternative, i.e., collecting original data on the members of the organizations.

[*Definition*] In tabulation, we use individual data to make a tabulation table. A subject described in the individual data as entity e has attributes that are represented by the variable X. We suppose that we have N entities, each of which has M variables. For an entity e_i represented by variables X_j^i, let the individual data E_i be represented as $\left(X_1^i, \cdots, X_j^i, \cdots, X_M^i\right)$. Tabulation tables $T_{\lambda \in \Lambda}$ are derived from a set of individual data E_i. Λ is the set of available tabulation table indexes.

[*Theory*] In tabulated data collection, when an attribute of individual data is added, the tabulated data size increases exponentially to maintain the equivalent space for storing individual data.

[*Proof*] To make the tabulation tables T_λ represent entity counts according to the values of variables, a variable X is categorical, or if it is numerical, it is translated into a categorical variable. Then, tabulation tables T_λ can be derived from the individual data set. They can be integrated into one tensor $\bar{T}$. $\bar{T}$ is a unified M-dimensional array to hold all frequencies in all tabulation tables T_λ. When the number of categories in a variable X_i is represented as $s(X_i)$, the size of tensor $s(\bar{T})$ is:

$$s(\bar{T}) = s(X_1) \cdots s(X_M) = \prod_{m=1}^{M} s(X_m).$$

However, the size of each individual data $s(E_i)$ is M, and if $\bar{E}$ is denoted for all individual data E_i, the total size $s(\bar{E})$ is:

$$s(\bar{E}) = N \cdot M.$$

This means that for equivalent information storage, tabulation table data size depends on the number of variables and the product of the number of categories for each of them, whereas the individual data size depends on the entity population and the number of variables.

Here, we suppose that a new variable is added to the individual data. Then, the difference in the size of tensor $\Delta s(\bar{T})$ is:

$$\Delta s(\bar{T}) = \prod_{m=1}^{M+1} s(X_m) - \prod_{m=1}^{M} s(X_m)$$

$$= \{s(X_{M+1}) - 1\} \prod_{m=1}^{M} s(X_m)$$

$$\geq 2^M.$$

(the minimum number of categories is 2)

The difference in the total size of individual data $\Delta s(\bar{E})$ is:

$$\Delta s(\bar{E}) = N \cdot (M+1) - N \cdot M = N.$$

This means that the size of the tensor increases by the product of the number of categories for the variables, whereas the size of the individual data increases in proportion to the entity population. If the number of variables becomes large, the tabulated data size becomes much larger than the individual data size.

Thus, under a limited data size, the individual data set is more powerful for analysis than the tabulated data derived from it.

4   Related Work

The National Center for Education Statistics (NCES), the statistical arm of the U.S. Department of Education (ED), provides the Integrated Postsecondary Education Data System (IPEDS) [7], which consists of several educational survey data sets from colleges and universities. When the survey was introduced in 1985–86, it collected data on a voluntary basis from all institutions whose primary purpose was to provide postsecondary education programs open to the general public. Today, it is a mandatory reporting system for institutions that participate in Title IV federal student financial aid, following a decision by Congress in 1992 [8]. As of 2016–17, the IPEDS survey components include institutional characteristics, admissions, fall enrollment, 12-month enrollment, student financial aid, completions, graduation rates, outcome measures, human resources, finances, and academic libraries [9].

IPEDS adopted an institutional-level data collection approach from the beginning, although a student-level data collection was proposed by the NCES and, following a feasibility study, endorsed by a political commission in 2005. The proposal faced considerable opposition, primarily from private and nonprofit colleges and universities, as well as from privacy groups on both the left and the right of the political spectrum [8].

IPEDS data have been utilized for purposes beyond their original ones. College Scorecard [10], operated by the ED, is a web-based tool for consumers to compare the cost and value of higher education institutions in the United States. It is built on the IPEDS data, which forms part of its database. Furthermore, IPEDS data are merged with a range of other federal and private-sector data sources so that researchers can demonstrate the value of higher education and evaluate the effectiveness of various policies and practices [11].

The Higher Education Statistics Agency (HESA) in the UK is an institute tasked with collecting, verifying, and disseminating data about higher education in the UK and is the designated data body for England [12]. It collects the *Aggregate Offshore* student record, the *Estates Management* record, the *Finance* record, the *General Medical Council Student Assessment* record,

the *Graduate Outcomes Contact Details* record, the *Graduate Outcomes Survey Results* record, the *Higher Education – Business and Community Interaction* record, the *Initial Teacher Training* record, the *Provider Profile* record, the *Staff* record, the *Student* record including *Unistats*, and the *Student Alternative* record.

Of the above collections, the *Staff*, *Student*, and *Student Alternative* records are individual-level data. The *Staff* record has been collected since 1994–95, when it was initially coded in a record format; it has been coded in XML format since 2012–13. The *Student* record has been collected since 1994–95, when it was initially coded in a record format; it has been coded in XML format since 2007–08.

The data collections are being used for purposes other than as an analytical basis for policy-making. Discover Uni [13] is an official source of information about higher education in the UK, designed to support prospective students in deciding whether, where, and what to study. It includes the *Student* record and the *Student Alternative* record as part of its collected data. It is operated by the UK higher education funding and regulatory bodies, i.e., the Department for the Economy in Northern Ireland, the Higher Education Funding Council for Wales, the Office for Students in England, and the Scottish Funding Council. Heidi Plus [14] is a higher education sector visualization and analytics tool for higher education professionals to gain the latest institutional and sector insights using interactive data visualizations and dynamic dashboards. It includes up to 12 years of HESA data, including *Students*, *Staff*, *Estates Management*, and *Graduate Outcome Records*. It is provided through a joint effort of the HESA data analytics team and Jisc Analytics Labs. Its latest community dashboard was built on the system for a higher education institution quality health check by collaborating with a team from the Quality Assurance Agency [15].

5   Conclusion and Future Work

In this paper, we reconsider the architecture of the information collected from higher education institutions by NIAD-QE and analyze its character. NIAD-QE periodically collects educational information, evidence-based documents, and self-assessment reports to enhance the quality of higher education, which are in a broad sense used to support university reform. These resources are a collection of organizational information for observing higher education institutions where, to collect facts about organizational members, the data are aggregated in a tabulated form, even though this is not as efficient a way of conveying the information content in a limited data size as the original individual data sets.

Based on this analytical work, we will design and implement a prototype for the next-generation information infrastructure for university reform in NIAD-QE for further study.

Acknowledgements

References

[1] National Institution for Academic Degrees and Quality Enhancement of Higher Education, Glossary of Quality Assurance in Japanese Higher Education, 5th ed. Tokyo: NIAD-QE, 2021.

[2] Standards for Establishment of Universities (Order of the Ministry of Education, Science and Culture No. 28 of 1956), https://www.japaneselawtranslation.go.jp/en/laws/view/3793, (Accessed 2022-04-20).

[3] National University Corporation Act (Act No. 112 of 2003), https://elaws.e-gov.go.jp/document?lawid=415AC0000000112, (Accessed 2022-04-20).

[4] School Education Act (Act No. 26 of 1947), https://elaws.e-gov.go.jp/document?lawid=322AC0000000026, (Accessed 2022-04-20).

[5] Enforcement Regulations for the School Education Law (Order of the Ministry of Education, Science and Culture No. 11 of 1947), https://elaws.e-gov.go.jp/document?lawid=322M40000080011, (Accessed 2022-04-20).

[6] Japanese College and University Portraits, https://jpcup.niad.ac.jp, (Accessed 2022-04-20).

[7] IPEDS homepage, https://nces.ed.gov/ipeds/, (Accessed 2022-04-20).

[8] E. S. Miller and J. M. Shedd, "The history and evolution of IPEDS," New Dir. Institutional Res., vol. 2019, no. 181, pp. 47-58, Mar. 2019.

[9] A. Aliyeva and C. Cody, "The history and origins of survey items for the Integrated Postsecondary Education Data System (2016–17 Update)," NPEC 2018-023, U.S. Department of Education. Washington, DC: National Postsecondary Education Cooperative, Mar. 2018.

[10] College Scorecard, https://collegescorecard.ed.gov, (Accessed 2022-04-20)

[11] R. Kelchen, "Merging data to facilitate analyses," New Dir. Institutional Res., vol. 2019, no. 181, pp. 59–72, Mar. 2019.

[12] HESA homepage, https://www.hesa.ac.uk/, (Accessed 2022-04-20).

[13] Discover Uni, https://discoveruni.gov.uk, (Accessed 2022-04-20).

[14] Heidi Plus, https://www.jisc.ac.uk/heidi-plus, (Accessed 2022-04-20).

[15] HESA, "Joint QAA, HESA, and Jisc new Community Dashboard is released | HESA," 23-Apr-2018, https://www.hesa.ac.uk/news/23-04-2018/qaa-community-dashboard, (Accessed 2022-04-20).