Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Proposing a Processing Distribution System for Cross-Organizational Use of Educational Data

Takahiro MORITA[a*], Yuko MURAKAMI[ab], Hidenobu WATANABE[ab] & Kouji NISHIMURA[ab]
[a]*Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*
[b]*Information Media Center, Hiroshima University, Japan*
*m221924@hiroshima-u.ac.jp

Abstract: Cross-organizational data utilization is crucial for fostering DX in education. In order to leverage data across multiple organizations, some research suggests platforms that enable data-using organizations to bring data and the programs deployed to a secure execution environment for processing. However, when it comes to using educational data, it is concerned about the risk of confidential information leakage by transferring the data outside of organization. In addition, to efficiently use data across organizations, each organization should perform common preprocessing on the data. In this paper, we propose a processing distribution system that can securely and efficiently collect educational data and utilize it across organizations.

Keywords: DX, Educational data, Cross-organizational use, Processing distribution, GakuNin

1. Introduction

To promote DX (digital transformation) in the field of education, it is necessary not only to analyze educational data held by each organization independently but also to enable cross-organizational analysis. Cross-organizational data utilization can be an opportunity to create new services, in addition to discovering and improving problems in education. Some research suggests platforms that allow the processing of data collected from various organizations in a secure execution environment while ensuring data security and preventing information leakage (NTT, 2022; Sakamoto et al., 2021). These platforms provide an execution environment that keeps data and programs secret from others and enable program execution on that environment. After executing the process, the execution environment is deleted, and the data generated after execution is encrypted and saved in secure storage to prevent information leakage. Organizations that hope to use the data can receive processed results after moving the data and the programs to an environment outside of organization.

When focusing on the utilization of educational data, it is considered difficult from the viewpoint of ensuring privacy to collect the raw educational data held by each organization and analyze it across the organization. Hence, processing platform without transferring the data outside of organization is required. It is also important to be able to verify that educational data is processed appropriately, ensuring transparency in the process.

We propose a processing distribution system that facilitates the cross-organizational utilization of educational data. This system distributes a processing environment for handling data to each organization using container-based virtualization technology. Container packages an application and execution environment together virtually, enables independent processing to run in all computing environments. The processing content in the container is also shared and distributed after prior testing to ensure its validity. This approach allows the organization to access a trusted processing environment and analyze raw educational data without needing to transfer the data outside the organization. Additionally, it enables data standardization and anonymization for cross-organizational analysis.

This paper outlines the requirements for collecting and utilizing educational data, then, introduces the proposed system along with a specific example of its application. Finally, it concludes with a summary.

2. Requirements for collection and utilization of educational data

Figure 1 shows the data utilization flow required to realize educational DX. There are two ways to utilize educational data: intra-organizational use and cross-organizational use. Intra-organizational use necessitates a system that enables concise data processing within an organization, without the need to transfer the data externally, and the results are utilized within the same organization (Figure 1-(a)). For cross-organizational use, data is typically anonymized and aggregated, but this can lead to reduced analysis accuracy. Therefore, it is essential to establish a mechanism to anonymize data (Figure 1-(b)) after preprocessing such as data standardization (MEXT, 2021) and analysis within each organization (Figure 1-(a)), followed by aggregation of the results (Figure 1-(c)). This allows the analysis accuracy to be maintained at a high level after data is aggregated. It is also crucial to validate the program before sharing and distributing, ensuring that multiple organizations can agree on how the data will be processed. Additionally, a seamless mechanism is required to connect intra-organizational use to cross-organizational use. By implementing a system equipped with these mechanisms, data utilization can be improved, allowing for better data utilization by refining the program based on knowledge gained from cross-organizational use.

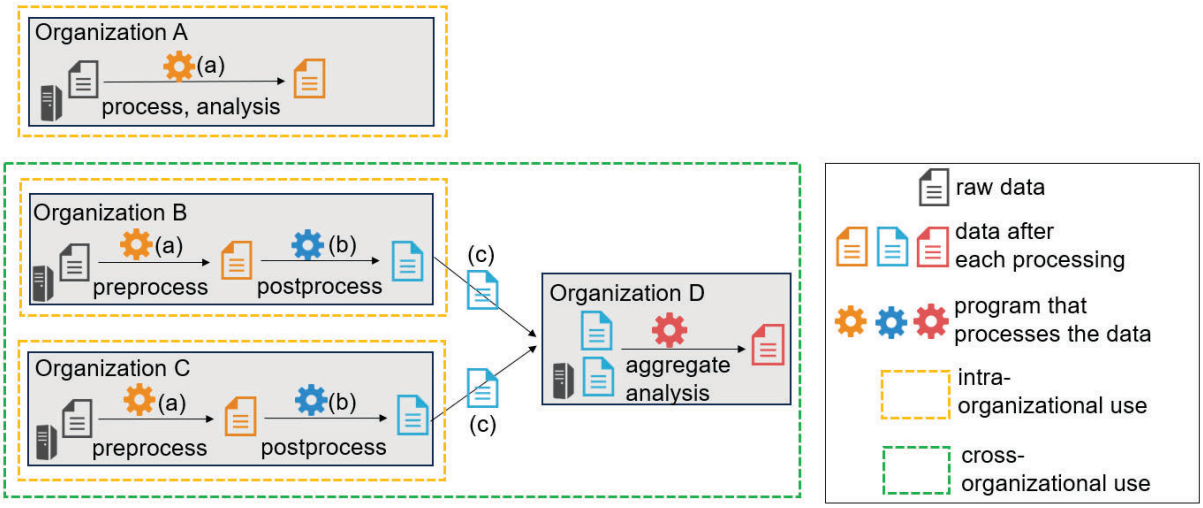

Figure 1. Data utilization flow required to realize educational DX.

3. Processing distribution system and a specific example of its utilization

To achieve the data collection and utilization described in Chapter 2, we propose a processing distribution system. Figure 2 shows an overview of the processing distribution system. Since containers are used in this system, the system can be used by installing Docker regardless of the server OS. The system operates on the premise that a group of servers executing the processing is connected through the Academic Access Management Federation in Japan (GakuNin, 2009), allowing container execution and other operations among organizations cooperating through GakuNin. External access from the container execution environment is restricted, and data export in a manner not desired by the data owner is prohibited.

The data processing workflow in this system commences with the container administrator creating a container image containing the application that will process the data. After testing verifies its validity, the container image is registered in the container registry.

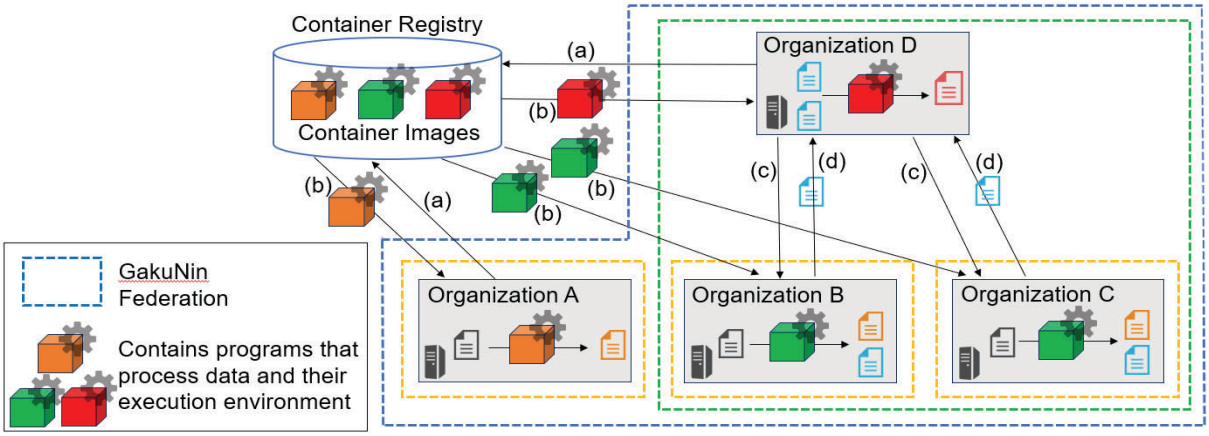

Figure 2. Overall view of processing distribution system.

After authentication with GakuNin, data owners then select and download the desired container image (Figure 2-(a)(b)) and execute the processing within their environment. This represents the intra-organizational use. For cross-organizational use, a member of an organization who wants to use the data obtains authorization to run the container for the cooperating organization and gives processing instructions to the system of each organization (Figure 2-(c)). Each organization's system then downloads the required container images according to the instructions. (Figure 2-(b)). The data resulting from the processing performed on each server is subsequently aggregated on the organization's server for further analysis (Figure 2-(d)).

An example application of this system is the utilization of school survey data. For instance, an educational institution intends to conduct a standardized survey across multiple schools. A web application containing the survey content is containerized and distributed to each school. Since the data to be utilized are responses to a questionnaire and the data format is the same, the processing of the data can also be containerized and each school can use the container for intra-organizational use. Moreover, by analyzing the survey results within each school and then performing anonymization and other processing using containers, the survey analysis results can be safely shared. This enables the comprehension and comparison of overall trends.

4. Conclusion

We have proposed a processing distribution system for the utilization of educational data across organizations. With this system, essential data can be efficiently processed within the organization's environment, without needing to transfer the data outside of the organization. Depending on differences in the content and format of data held by organizations, it may be difficult to use data only with container processing. However, this system enables uniform data collection and analysis across organizations by defining anonymization policies and data format rules in advance.

References

NTT (2022). Next-generation Data Hub for Secure and Convenient Data Utilization across Organizational Boundaries. Retrieved August 17, 2023, from https://www.rd.ntt/e/research/JN202202_17186.html

Sakamoto, H., Ishida, K., Kato, T., & Inagaki, Y. (2021). “Data value Co-creation Platform” which shares organization’s data among other organizations, and enables multiple organizations to increase the value of data [in Japanese]. DC, 2021-DC-120(10), 1-8.

MEXT (2021). Discussion Paper on the Utilization of Educational Data [in Japanese]. Retrieved August 17, 2023, from https://www.mext.go.jp/content/20210331-mxt_syoto01-000013887_1.pdf

GakuNin (2009). About GakuNin. Retrieved August 17, 2023, from https://www.gakunin.jp/en