Nationwide Research Data Management Service of Japan in the Open Science Era

Yusuke Komiyama
Digital Content and Media Sciences Research Division,
National Institute of Informatics
Chiyoda Ward, Tokyo, JAPAN
komiyama@nii.ac.jp

Kazutsuna Yamaji
Digital Content and Media Sciences Research Division,
National Institute of Informatics
Chiyoda Ward, Tokyo, JAPAN
yamaji@nii.ac.jp

***Abstract*—Recently, in Japan, there has been a great need in universities and research institutions to archive their research data for ten years because they need to maintain the reproducibility of data for ensuring research integrity and the promotion of open science. However, a research data management (RDM) service does not exist in Japan. Therefore, at the National Institute of Informatics (NII), we developed the next-generation RDM on a national scale by using the open science framework (OSF). We combined the RDM with the existing NII worldwide cyberinfrastructure services (SINET and GakuNin) and OSF add-ons for external Cloud services and institutional repositories. Finally, we displayed the first closed trial for nationwide RDM services in seven universities and a national research institute for research integrity; this helped promote open science in Japan in the fourth quarter of fiscal year 2016. NII and participant organizations obtained some RDM operating know-how and discovered new issues by follow-up meetings. The source code is available on GitHub at (http://doi.org/10.5281/zenodo.546481);(http://doi.org/10.5281/zenodo.546480).**

Keywords—RDM; research data management service; research data archive; research support service; research integrity; open science;

I. INTRODUCTION

First, academic organizations are promoting platform implementations for the open science by all over the world. The statement of the committee on open science originated from the declaration made in the Tsukuba Communiqué by the G7 science and technology ministers' meeting in Tsukuba, Ibaraki, in 2016 [1].

On the other hand, in Japanese academia, the data archive for research integrity is an important subject, and we are focusing on the open science as next step. The committee on open science of the Science Council of Japan (SCJ) made "recommendations concerning an approach to open science that will contribute to open innovation" [2]. This statement describes the concept of research data cyberinfrastructure for open science to the academia in Japan, and it defines three factors of cyberinfrastructures, which includes the management infrastructure, publication infrastructure, and discovery infrastructure for research data. Besides, all universities and research institutions need to archive research data from each laboratory for ten years because this will maintain the research integrity. Research data archives will be required for projects of the Grants-in-Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) or the Japan Society for the Promotion of Science (JSPS). Besides this, the Japan Science and Technology agency (JST) required the submission of a data management plan for the adoption of researchers in the CREST and SAKIGAKE grant programs from the 2016 fiscal year, on an experimental basis.

Therefore, Japanese academic organization needs to prepare to exist research data management (RDM) of a compatibility nationally, in the Occident, some organization are servicing advanced RDM; in the United States, the Center for Open Science (COS), a non-profit organization, created the Open Science Framework (OSF), which is a research data management service [3]–[5]; in Europe, the Collaborative European Data (EUDAT) infrastructure is provided for the European science community under the European Open Science Cloud [6]–[9].

However, the trusted service of RDM does not exist in Japan as yet. Therefore, NII proposes to develop the next-generation RDM and provide sustainable services for the cyberscience infrastructure center and library at the Japanese university. This study aims to maintain research integrity and promote open access.

II. SYSTEM DEVELOPMENT

A. Open Science Framework as a core system

COS publishes OSF source code as open source software on GitHub, and they have developed OSF using Python, which is the feature of OSF expansibility. The OSF application uses the Central Authentication Service (CAS) for the

978-1-5386-0621-6/17 $31.00 © 2017 IEEE
DOI 10.1109/IIAI-AAI.2017.144

authentication, and it has interoperability with Shibboleth. We adopt OSF as the core system software in the new RDM service, fork the original OSF source cord, and remake it into the new RDM service for Japanese researchers.

B. GakuNin as a federation for authentication

GakuNin is the nationwide authentication federation for Japanese universities; it is a useful service solution for single sign-on for the application developer and end users because GakuNin hosts the Shibboleth-based solutions having identity providers, service providers, and discovery services. We use GakuNin for the new RDM service [10]. We developed a connecting module between Shibboleth and CAS for OSF—the GakuNin connection [11].

C. SINET as high-speed backbone networks for science

Science Information Network (SINET) is the high-speed backbone network for more than 800 Japanese universities and research institutions [12]. The Cloud connection service connects an L2-VPN between an on-premise system in the university and the Cloud data center via SINET. We show the concept system design of the NII-RDM service using the public Cloud AWS. Figure 1 shows an imaginary private Cloud with the GakuNin federation.

D. OSF add-on development for external Cloud services

The university cyberinfrastructure centers are using public/private Cloud storages, such as Amazon Web Services (AWS) S3, Microsoft Azure Blob Storage, ownCloud, OpenStack Swift. We propose central aggregation to the Web application and database, and we desire to distribute the storage into each university and research institution. The original OSF software does not have the add-ons necessary for the external services of the Japanese university. Therefore, first, we investigated the system component relationship of the original OSF add-ons in each Web architecture layer for the OSF add-on development. Figure 2 shows the relationship diagram of OpenStack Swift. Next, we included two add-ons in the external Cloud storages for Microsoft Azure Blob storage (public Cloud) and OpenStack Swift (private Cloud).

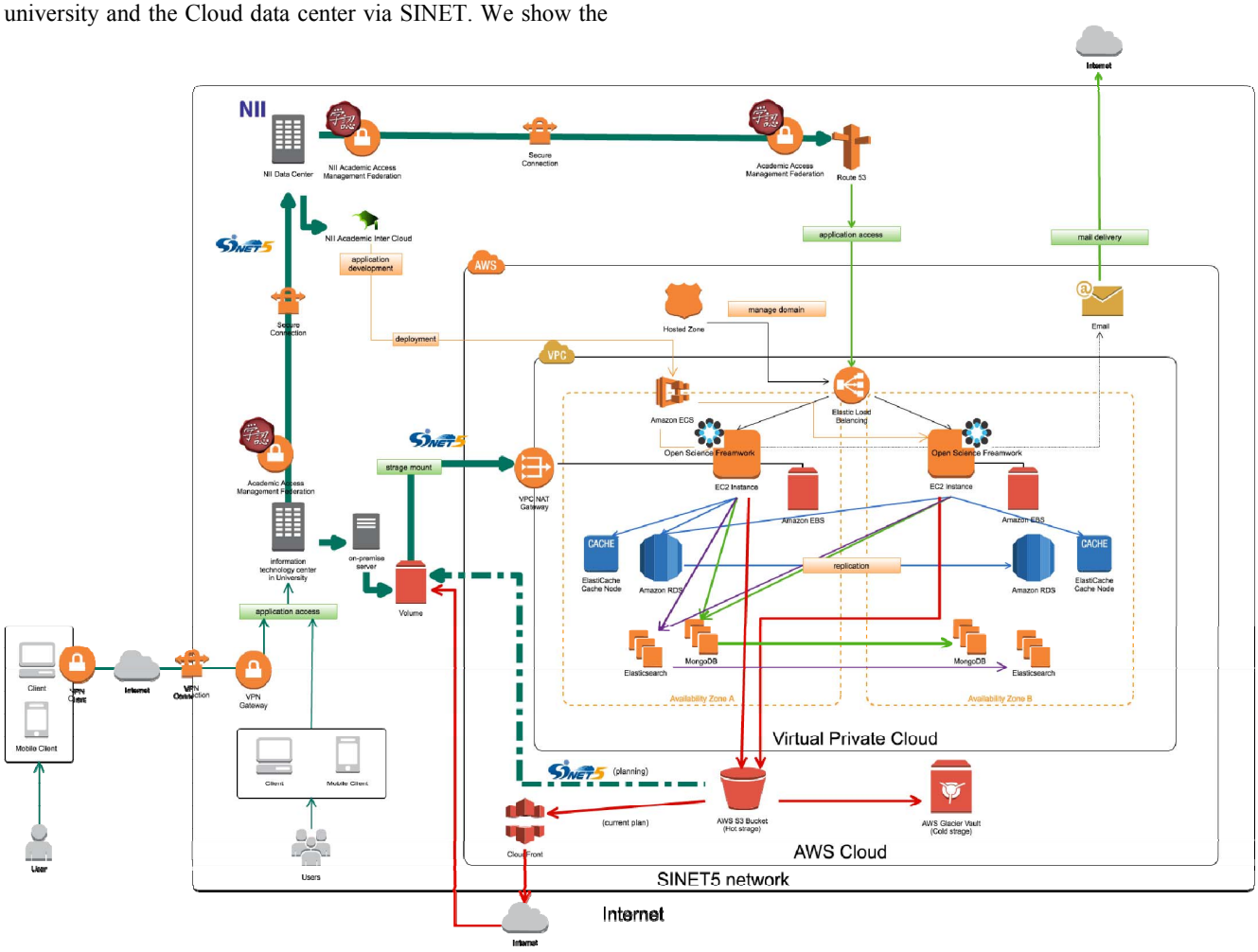

Fig. 1. Concept system design of NII-RDM service using AWS on the SINET with GakuNin federation. Bold green arrows are SINET lines, blackish red logos are single sign on by GakuNin.

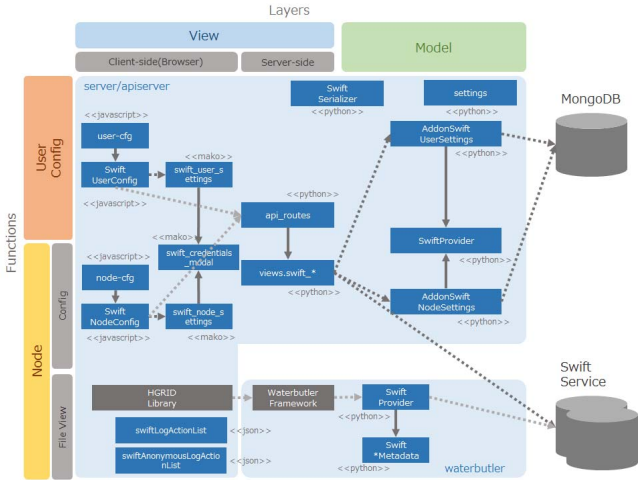

Fig. 2. System component relation of OSF add-ons for OpenStack Swift in each Web architecture layers.

E. OSF add-on development for institutional repository

The original OSF includes the Dataverse add-on for external institutional repository (IR) services [13], but major Japanese university libraries and research institutions are adopting the WEKO widely. First, we customized some WEKO repositories for the OAuth applied to OSF. Figure 3 shows the method for implementing the system components and protocol workflows. This method improves the outer service side for the WEKO provider plugins using PHP; the plugins observe Simple Web-service Offering Repository Deposit protocols [14].

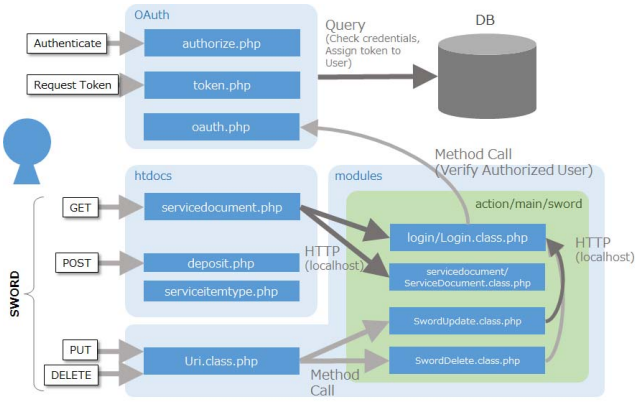

Fig. 3. System component and protocol workflow of OAuth applied OSF.

In addition, university libraries are operating the JAIRO Cloud with WEKO, which is a service as a service (SaaS) of the IR launched by the National Institute of Informatics of Research Organization of Information and Systems (ROIS-NII) [15]. we made the OSF WEKO add-on experimentally and connected it to the customized WEKO repository on a virtual machine. Figure 4 shows the system component relationship of OSF add-ons for WEKO in each Web architecture layer.

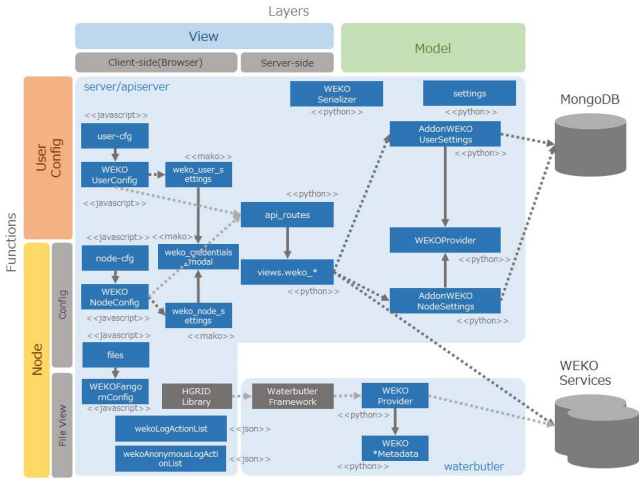

Fig. 4. System component relation of OSF addons for WEKO in each Web architecture layers.

F. Computing resources and system environments

Academic inter cloud (AIC) is a managed service of Cloud computing resource; it is of bare-metal type and is used for scientific research by NII in Japan [16]. We prepared virtual machines on the AIC instances (Intel Xeon E5-2670 2.60GHz, 8Core, 96 GB memory, and 1.8TB × 1, 1TB × 4 storage); we selected the instance type *r3.8xlarge* (32 virtual CPUs, 244 GiB memory, and 320 GB × 2 SSD storage) in AWS Elastic Compute Cloud (EC2) as the public Cloud services. Its detailed specifications are shown in Table 1 and Table 2.

TABLE I. CLOUD COMPUTING RESOURCES OF NII-RDM ON AIC

Academic Inter Cloud (hypervisor)	
Physical product name	Dell PowerEdge C6220
Operating system	CentOS 6
Hypervisor type	Kernel-based Virtual Machine (KVM)
IaaS platform	OpenStack
CPU	Intel (R) Xeon (R) CPU E5-2670 2.60GHz, 8 Core 16 Thread (× 2CPU)
Memory	96GB 8GBx8 + 4GB × 8/2R/1600MHz/RDIMM
Disk storage	RAID10 1.8TB × 1 Volume 1TB × 4 SAS 2.5inch HDD
Network interface card	10GbE NIC 2
Academic Inter Cloud (guest OS)	
Operating system	Ubuntu 14.04.4 LTS (CNU/Linux 3.13.0-91-generic_x86_64)
Memory	16GiB
Disk storage	100GB

TABLE II. CLOUD COMPUTING RESOURCES OS NII-RDM ON AWS

Amazon Web Services (Elastic Compute Cloud)	
Instance type	r3.8xlarge
vCPU	32
Memory	244GiB
SSD storage	2 × 320GB

III. Results

We designed the prototype model of the nationwide RDM service using the existing academic cyberinfrastructures in Japan and named the system NII-RDM. We combined the customized OSF on the SINET, GakuNin, AIC, and AWS, and developed the OSF add-ons for three extra services that included the add-on of Microsoft Azure Blob storage, OpenStack Swift, and WEKO by using standard APIs and protocols. Also, we separated the system components of the original OSF into the Docker containers and reconstructed the Docker compose for the NII-RDM. Institutional IT administrators could connect their managed storage and IR to the NII-RDM service easily. Figure 5(a) shows the storage connection methods using OSF; figure 5(b) shows three new OSF add-ons on NII-RDM; figure 5(c) shows the prototype NII-RDM portal screen. Also, we prepared the technical document and the user manual for the NII-RDM and distributed the related source code for domestic engineers. We shortened the time required for completion and thereby reduced costs. For the next five years, we have proposed a system federation ranging from NII-RDM to the nationwide next-generation IR for research data sharing. Besides, we detected original English messages in the system component of the OSF source codes for developing the multilingual selector functions. Therefore, we concluded that the university staff members and students needed a multilingual selector function on the NII-RDM in Japan.

Finally, we held the first NII-RDM closed trial with the eight universities/institutions for three weeks from February to March 2017. We collected feedback about NII-RDM's trial service from IT faculty members and managers placed in universities/institutions; this information was collected on the Wiki. In this trial, we paid particular attention to multifactor certifications for security using a firewall control and GakuNin authentication.

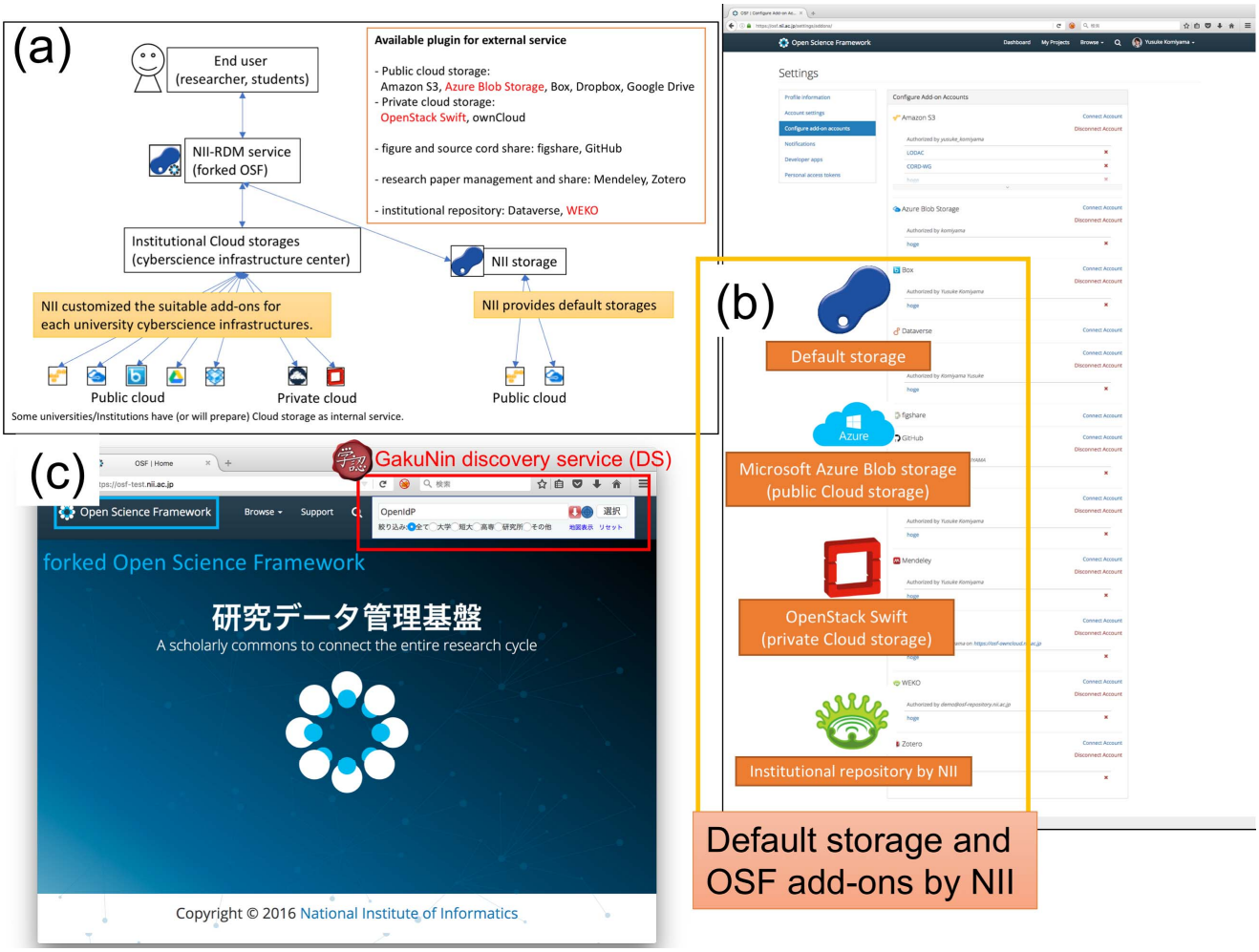

Fig. 5. (a) Storage connection methods using OSF for Japanese university/institutions. (b) NII enhanced OSF software as NII-RDM about default storage and three add-ons. (c) the GakuNin discovery service for authentication of single sign on; and the customized core service.

IV. Conclusion

In this paper, we proposed the concept of a novel nationwide service to manage research data for Japanese academia, and we reported the results of the first trial. NII will provide nationwide RDM services to universities and institutions for ensuring research integrity and open access. NII-RDM can quickly provide a corroborative research environment for the group. In March 2017, we tested the service in a trial environment with eight institutions and collected the feedback from university IT managers about the system development. We are going to repeat a development and a user test continuously and improving the RDM service with Japanese academia. We established the NII Research Center for Open Science (RCOS) and a data platform in Japan; this research center will play a central role for research data integration. Their aims and objectives are to develop and promote the next-generation RDM service and IR systems for cyberscience infrastructure centers and libraries of universities/institutions for promoting research integrity and open science in Japan.

Availability

We released source codes about customized OSF (NII-RDM) and new add-ons under the Apache License version 2.0 on the following URLs. http://doi.org/10.5281/zenodo.546481; http://doi.org/10.5281/zenodo.546480.

Acknowledgment

The research data management service was provided by Research Center for Open Science and Data Platform (National Institute of Informatics).

References

[1] CabinetOfficeGovermentofJapan, "Tsukuba Communiqué," 2016.

[2] CommitteeOnOpenScience_ScienceCouncilofJapan, "Recommendations Concerning an Approach to Open Science That Will Contribute to Open Innovation," 2016.

[3] E. C. McKiernan, P. E. Bourne, C. T. Brown, S. Buck, A. Kenall, J. Lin, D. McDougall, B. A. Nosek, K. Ram, C. K. Soderberg, J. R. Spies, K. Thaney, A. Updegrove, K. H. Woo, and T. Yarkoni, "How open science helps researchers succeed.," *Elife*, vol. 5, no. JULY, Jul. 2016.

[4] M. McNutt, K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, and J. L. King, "RESEARCH INTEGRITY. Liberating field science samples and data.," *Science*, vol. 351, no. 6277, pp. 1024–6, Mar. 2016.

[5] B. A. Nosek, "Center for Open Science: Strategic Plan," Open Science Framework, 2017.

[6] European Commission, "European Open Science Cloud," *Nat. Genet.*, vol. 48, no. 8, pp. 821–821, Jul. 2016.

[7] S. B. Ardestani, C. J. Hakansson, E. Laure, I. Livenson, P. Stranak, E. Dima, D. Blommesteijn, and M. van de Sanden, "B2SHARE: An Open eScience Data Sharing Platform," in *2015 IEEE 11th International Conference on e-Science*, 2015, pp. 448–453.

[8] EUDAT, LIBER, OpenAIRE, EGI, and GEANT, "Position Paper: European Open Science Cloud for Research," Jan. 2015.

[9] A. S. Memon, J. Jensen, A. Cernivec, K. Benedyczak, and M. Riedel, "Federated Authentication and Credential Translation in the EUDAT Collaborative Data Infrastructure," in *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, 2014, pp. 726–731.

[10] K. Yamaji, T. Kataoka, M. Nakamura, T. Orawiwattanakul, and N. Sonehara, "Attribute Aggregating System for Shibboleth Based Access Management Federation," in *2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*, 2010, pp. 281–284.

[11] H. Naito, S. Kajita, Y. Hirano, and K. Mase, "Multiple-Tiered Security Hierarchy for Web Applications Using Central Authentication and Authorization Service," in *2007 International Symposium on Applications and the Internet Workshops*, 2007, pp. 27–27.

[12] T. Kurimoto, S. Urushidani, H. Yamada, K. Yamanaka, M. Nakamura, S. Abe, K. Fukuda, M. Koibuchi, Y. Ji, H. Takakura, and S. Yamada, "A fully meshed backbone network for data-intensive sciences and SDN services," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2016, pp. 909–911.

[13] G. King, "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing," *Sociol. Methods Res.*, vol. 36, no. 2, pp. 173–199, Nov. 2007.

[14] G. Mosweunyane and L. A. Carr, "Direct desktop-repository deposits with SWORD," in *2014 IST-Africa Conference Proceedings*, 2014, pp. 1–8.

[15] K. Yamaji, T. Aoyama, M. Furukawa, and T. Yamada, "Development and Deployment of the Open Access Repository and Its Application to the Open Educational Recourses," Springer, Cham, 2016, pp. 395–403.

[16] S. Yokoyama, A. Matsumoto, and N. Yoshioka, "Network traffic optimization architecture for scalability in academic inter-cloud computing environments," in *Proceedings of the 2nd International Workshop on Hot Topics in Cloud service Scalability - HotTopiCS '14*, 2014, pp. 1–6.