Contribution ID: **69**                                                      Type: **Oral Presentation**

# Open Platform for Academic Humanities Data

*Wednesday, 16 March 2016 14:00 (30 minutes)*

Universities are major stakeholders of academic data. Kyoto University, since its foundation in 1897, has collected, created and accumulated numerous and various materials, data and knowledge as its academic resources, and it has developed databases for researchers to access these resources, i.e., KULINE (the university OPAC operated by the library), KURENAI (the university repository developed by the library), KURRA (the university research archives developed by the museum), University Open Course Ware operated by the Academic Center for Computing and Media Studies, and various databases developed by research institutes/centers. These databases include resources of research processes from data collections (original materials, observation data and experimental data etc.) to publications (papers, books etc.) and archives. However, as each database is independent and heterogeneous system, it is difficult to carry out even simple searches such as finding original experimental data related to the paper. Obviously, we cannot use such isolated databases for advanced research usages to discover hints and/or create new knowledge.

Kyoto University has just launched a new project to develop an innovative database platform adapted to Cloud and Big Data environments to accumulate and link its academic resources, and offered the platform as advanced research utilities. This platform will comprise three sublayers. The fist layer is "Open Data Layer" to accumulate and open heterogeneous data. This layer uses RDF (Resource Description Framework) that can describe data of different structures by uniform way. For example, this layer can accumulate thesauri (tree structure), bibliographic catalogues (table structure) and documents (XML) simultaneously. The second layer is "Data Link Layer." Academic data, especially humanities' data, are ambiguous (i.e., a term "book" in a database may be expressed as "⬚" in the other database, "purple" may be the same notion of "⬚," and "cat" and "dog" may be the same category because of they are subordinate concept of "mammal"). This layer uses ontology techniques such as RDFS and OWL to link ambiguous notions and/or vocabularies and to create "Academic Big Data." Academic Big Data comprise small fragments of heterogeneous data, which will form complex structures. This is the different feature from ordinary big data comprising simple structure data from sensors and IoT devices. The third layer is "Application Layer." As Academic Big Data is too huge and complicate for researchers to retrieve, categorize and analyze by hands, and then applications to support these processes are necessary. The new project will develop some utilities, i.e., to estimate subjects of contents by natural language techniques, to categorize huge data sets by deep learning techniques, and to organize data according to topological spatiotemporal expressions. This platform will also provide APIs to create mashup applications easily.

This presentation will describe overview and state of progress about our project to promote advanced usages of academic resources of Kyoto University as "linked open data" on Cloud environment.

**Primary author:**   Prof. SHOICHIRO, HARA (Center for Integrated Area Studies, Kyoto University)

**Presenter:**   Prof. SHOICHIRO, HARA (Center for Integrated Area Studies, Kyoto University)

**Session Classification:**   Humanities, Arts, and Social Sciences Session I

**Track Classification:**   Data Management