
Constructing Linked Knowledge around Southeast Asian Studies

Akihiro Kameda

kameda@cseas.kyoto-u.ac.jp
Center for Integrated Area Studies
Kyoto University, Japan

Shoichiro Hara

shara@cseas.kyoto-u.ac.jp
Center for Integrated Area Studies
Kyoto University, Japan

To understand an academic humanities paper, it is crucial to understand each element in the paper (person, place, mention of another paper, etc.). To understand each element in the paper, we have to refer to knowledge outside of the paper; in other words, linked knowledge is important. Linked Open Data (LOD) (Heather and Bizer, 2011), which is emerging from research in the Semantic Web, provides the way to represent and share such linked knowledge.

We prepared texts of "Japanese Journal of Southeast Asian Studies" as a core dataset, in order to attempt to link words or documents in the core dataset to external resources such as [DBpedia](#) (also available [in Japanese](#)) or the [National Diet Library](#) in Japan.

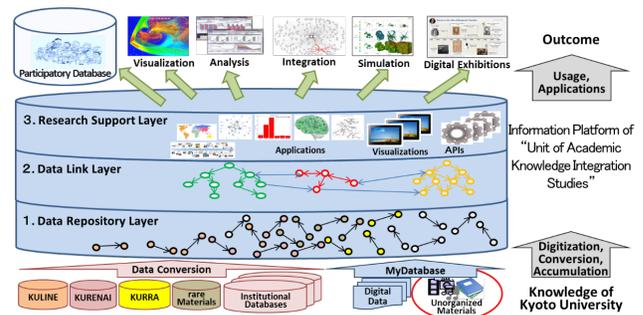
Though LOD has been growing rapidly (Schmachtenberg, 2014), it is difficult to cover specific knowledge in each academic paper. Therefore, the publication of LOD is also an important effort to represent knowledge networks in academic humanities papers.

The Center for Integrated Area Studies, Kyoto University (CIAS) has developed two information tools, named MyDatabase (MyDB) and Resource Sharing System (RSS), to solve these difficulties. The main component of MyDB is a database builder, allowing humanities researchers to construct and revise databases without expert knowledge. MyDB stores metadata and accepts any vocabulary of metadata, including nonstandard vocabularies. This enables humanities researchers to use their own metadata vocabulary according to their own purposes. On the

other hand, those metadata varieties make the integration processes difficult. RSS was developed to integrate heterogeneous databases on the Internet, and to provide users with a uniform interface to retrieve databases seamlessly in one operation. Thus, MyDB and RSS have helped accelerate open data in the humanities. However, there are still two problems to solve, especially in the case of RSS: limited coverage of databases and initial costs of integration. First, for example, Kyoto University released KULINE (OPAC), KURENAI (repository), KURRA (archive), Open Course Ware and various databases developed by each research institute in the university, but RSS does not integrate these databases. Second, it is time consuming to integrate new databases into RSS and impossible to trace links automatically. As such, for now, RSS is not the appropriate tool to discover hints and/or create new knowledge.

To overcome these drawbacks, a new project has been launched to develop an innovative information platform for open humanities data. This platform comprises three sublayers. The first layer is "Open Data Layer" which accumulates heterogeneous metadata. This layer uses RDF to describe data of different structures. The second layer is "Data Link Layer." This layer uses ontology techniques such as RDFS and OWL to link ambiguous (uncontrolled) vocabularies and emerge "humanities big data." The third layer is "Application Layer." As big data in the humanities is too huge and complicated to retrieve, categorize, and analyze by hand, this layer provides utilities to process big data. This platform will prepare for APIs to help mashup applications. We expect the platform to reconstruct a knowledgebase from heterogeneous databases, which is used to construct meaningful chunks from scattered data.

Thus, humanities Linked Open Data has been developed, and the "Japanese Journal of Southeast Asian Studies" dataset can be linked to that LOD. This linked knowledge can then help readers from other domains.



Bibliography

Heath, T. and Bizer, Ch. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers.

Schmachtenberg, M. et al. (2014). *Linking Open Data cloud diagram* <http://lod-cloud.net/>