

Historical Big Data - The challenge towards integrating historical document records -

*Mika Ichino¹, Junpei Hirano⁴, Kooiti Masuda³, Asanobu Kitamoto^{2,1}, Kenjiro Shou⁵

1. Center for Open Data in the Humanities, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 2. National Institute of Informatics, 3. Faculty of Urban Environmental Sciences, Tokyo Metropolitan University, 4. Faculty of Liberal Arts, Teikyo University, 5. Department of Civil Engineering, Nagoya Institute of Technology

Numerous historical documents in Japan detail a variety of events, such as earthquakes, tsunamis, the weather, stars, cherry blossom flowerings, aurora and so on. These records have been used for research in a number of fields ranging from scientific research including seismology, climatology and astronomy, to humanities, social science fields, and historical studies. Also, researchers sometimes extract data from the same documents. For example, the Ishikawa diary, written by a farmer family since 1720 in suburban Tokyo, not only includes daily weather conditions, but also others records such as meteorite and earthquake events. However, individual research groups have extracted and digitized the necessary parts to them. Over the years, the research studies using these records have been progressing in individual fields. As well as in historical studies, each researcher studies individual materials in subdivided areas. This study proposes the establishment of 'historical situation records', which is data such as the above historical data, and is also a concept for the data. Defining this schema will be the ability to share this data and share experiences using historical situation records. In addition to this will be developing structured data of the historical situation records that have been used in natural sciences, humanities and social sciences studies to integrate and analyze them, the "historical big data", to research for the temporal and spatial association between natural and human. For example paleo-climate and agriculture, disaster, population, and so on. It is possible this might lead to the creation of new knowledge. Recently, the digitization of historical textbooks and texts including historical situation records has progressed with the development of information technology, and it has been getting easier to use historical materials. Nonetheless, unlike modern big data, the data of historical situation records lacks structure for integrating and analyzing.

As one of our assignments, we are building an inter-disciplinary network for that purpose, and are holding the 6th CODH seminar "History big data ~ challenging toward integration and analysis of historical document records ~" on 12th March, 2018. A system to share the variety of information, experience and knowledge obtained from each individual research field is currently being developed.

Keywords: Historical document records, Data sharing, Integration and analysis, Paleoclimate