

Japan Data Repository Network (JDARN): Community-based activities for improving the trustworthiness of data repositories

*Asanobu Kitamoto^{1,2,4}, Hiroko Kinutani³

1. National Institute of Informatics, 2. ROIS-DS Center for Open Data in the Humanities, 3. EDITORIA, The University of Tokyo, 4. SOKENDAI

1. JDARN: establishment and brief history

Japan Data Repository Network (JDARN) is about community-based activities for sharing recent trends in the world and for improving trustworthiness of Japanese data repositories. Its origin is in “Project for assigning DOI on research data” hosted by Japan Link Center (JaLC) between October 2014 through September 2015. This project became a unique forum in which the experts of research data gathered across disciplines for the first time in Japan. As a follow-up activity, Research Data Utilization Forum (RDUF) was established in June 2016, and a few special interest groups (SIG) were later proposed. One of them is “networking domain repository stakeholders in Japan” SIG on October 2017. Then, on October 2018, it was upgraded to “Japan Data Repository Network” to represent our intention to expand the community to more disciplines and more stakeholders.

The purpose of JDARN is to share recent trends of the data repository, and in particular, it focuses on the issue of trustworthiness of data repositories. Trustworthiness plays an important role on decision making when data producers choose one of external services to deposit their data. One of the criteria to show the trustworthiness is CoreTrustSeal (CTS). CTS is one of international certifications on data repositories, and as of February 2018, over 140 data repositories have received certification, but in Japan the number of certification is still few. To understand the reason of delayed adoption of CTS in Japan, we held a seminar "Trustworthy Data Repositories - Forum for Sharing Practical Information about CoreTrustSeal Certification -" on December 2017, so that major data repositories in Japan could try their self-assessment according to the requirements of CTS. As a result, we realized that self-assessment by CTS is difficult unless we understand the fundamental concepts behind CTS. Hence the SIG started to create documents for understanding CTS, and this grew into the main activity of the SIG. Then, after intensive discussion, the focus of the activity has shifted to creating the data repository guideline that takes consideration of not only CTS but also data utilization.

2. Data Repository Guideline

Data repository guideline we are creating now is referring to 16 requirements of CTS, but is not its direct translation and has a unique structure proposed by JDARN. Rethinking of CTS has started by the item-based organization of CTS proposed by Mr. Shigeru Yatsuzuka at National Bioscience Database Center (NBDC). In the review process of CTS, the creation and publication of many types of documents is considered as an evidence of transparency. From the analysis of actual applications accepted by CTS and the type of documents mentioned therein, we can identify next actions to prepare necessary documents for CTS. In other words, we may be able to create an easier-to-understand guideline by converting abstract descriptions of CTS into concrete entities such as people and documents.

However, we also realized that itemizing people involved in data repositories is more difficult than itemizing documents. What kind of jobs do we need in data repositories, and who should be in charge of

them? We also have a problem of naming; namely how those experts should be called. People proposed many new names for data experts, such as data librarian, data curator, data scientist, data engineer, and so on, but the actual meaning of those words are different from person to person. We need to organize the concept of jobs, and demonstrate their long-term career paths; otherwise, open science based on data repositories has uncertain future. We still do not have a solid model for this issue, and we are still in discussion.

3. Future directions

Since the establishment of JDARN, we have had active discussions in meetings held roughly every month. We believe that, as more data repositories join the discussion, we have more data repositories in Japan which are high quality, show more presence in the world, and has greater value as the infrastructure of open science. To realize this goal, data repositories should be considered as indispensable entity in the research. The main focus of CTS is to improve the trustworthiness and sustainability of data repositories as the container of data, but we also need experts for utilizing the content of data such as data integration, data analysis, data visualization and societal impact. A single organization is difficult to take care of all the tasks, so collaboration among data repositories will also be an important issue, and this is where network of data repositories can play an important role.

Keywords: Japan Data Repository Network, Data repository, Trustworthiness , CoreTrustSeal, Guideline, Data expert