ISSN 2432-1249
SPARC Japan NewsLetter
NO. 25 Mar. 2015

The 4th SPARC Japan Seminar 2014

"What Should We Do to Expand Green Content?"

Monday, March 9, 2015: National Institute of Informatics
12th floor conference room (Attendees: 68)

Opening access to the diverse academic resources resulting from scholarly research is demanded not only from an academic standpoint but as a response to the needs of society. Looking at the current situation regarding academic resources in Japan, however, there are still large gaps among different research fields and different institutions and organizations in their awareness of and motivation to achieve open access, as well as the state of its accomplishment. Moreover, in many institutions and organizations, the academic resources they generate are spread among multiple systems, making it urgent to devise a comprehensive management structure model for systematic promotion of open access.
Currently the majority of content archived in institutional repositories consists of scholarly articles, whereas actual research results include much more diverse materials including research data and specimens. In this seminar, the academic content made public and disseminated by scholarly research institutions in institutional repositories and by other means is first of all redefined as "green content." On this basis, we then tried to find specific ways of managing academic resources, including building collections and encouraging use, with the scope of open access expanded to research data (other than scholarly articles), metadata for museum materials, visual data, and the like. We hope the seminar was a first step toward coming to a common awareness of what open access to academic resources can mean in the future.
A summary of the seminar is given below. See the SPARC Japan website (http://www.nii.ac.jp/sparc/event/2014/20150309.html) for handouts and other details.

Presentations

A Path to Data Management by Libraries

Yasuyuki Minamiyama
(National Institute of Polar Research)

After "open access" was defined in the Budapest Open Access Initiative in 2002, activities in this area have spread widely, to the point where today in 2015 we can say there is no one in academic publishing who is unaware of this concept. In recent years, moreover, it is taking on a broader meaning under the keyword "open science," the expectation being that it will not stop with open access to scholarly articles but will come to include openness of the underlying data and research process. In this seminar, I would like to try to connect research data, museum materials, and other such "data" in line with the open science concept, with its hopes for innovation through knowledge recycling and reuse from open access to data and research processes, and for encouragement of collaboration with other industries.
Why is data so important in the first place? The concept of "data-centered science" has been proposed recently, emerging from the desire to make more effective use of data, which is both the basis and result of research. In data-centered science, scholarly articles are also treated as data, with the hope being that a cycle will arise by which new data is created based on collected large-scale, complex data (data-driven research). At the same time, the importance of the systematic management of data is coming to be recognized anew, in relation to the issue of research data governance, as an approach to dealing with research misconduct and achieving research transparency.
Overseas, funding agencies like the DCC (Digital Curation Centre) and RDA (Research Data Alliance), having been quick to recognize the significance of preserving and making available data like the above, are providing support for release of such data. Prompted by these moves, university libraries have also begun initiatives in this area. The University of Edinburgh in the UK and Purdue University in the US are among institutions that have launched data repositories and are actively supporting data availability. There

are moves in Japan as well by the community of researchers, funding agencies, and museums to make data available. Since around December 2014, there have been very active discussions, for example, in the Cabinet Office, the National Diet Library, and the Science Council of Japan. Against this background, we need to consider the role to be played by university libraries and specific ways of collaborating with other communities.

Initiatives in a University Museum for Open Access to Information on Academic Materials

Shunsuke Yamashita
(Kyoto University Unit of Synergetic Studies for Space)

Initiatives for Open Access and Linked Open Data are aimed at efficient sharing and exploitation of information, while also holding promise for the accumulation of additional information on the resources. Compared, however, to completed articles and journals, or information produced as an integral part of actual public services and the like, studies have not yet progressed on the more "labor-intensive" activities of producing information on academic materials such as those in museum collections and academic resource archives. I am involved in creating academic resource archives in the Kyoto University Museum. I described how, in a university museum, the research processes are closely related to the processes of creating and making academic specimens (the university museum is central to object-based research), and introduced the current reality of the around 2.6 million academic specimens in the Kyoto University Museum. In making academic materials information open, it is important to maintain and ensure the relationship "article→specimen, specimen → article"; and in the case of taxonomy, for example, when authorizing the new scientific name (publishing in a scholarly article), it is standard practice (Code of Nomenclature) to include information about voucher specimens also in the article. The Research Resource Archive currently being developed by Kyoto University is systematically archiving various materials collected or created in the education and research processes at Kyoto University, with the aim of making use of these materials as resources in new education and research. This project is collecting such materials as films and diaries that were outside the scope of existing repositories in the university, and is making them available vie Peek, the Kyoto University Digital Archive System. Using as an example the Hotta (Mitsuru) Movie Collection (from 1960 to ca. 1982), I showed the importance of associating plant specimens with film materials. In this regard, I noted that for making archives public it will be necessary to find solutions to such issues as assigning unique identifiers and dealing with the hierarchical nature of the archive materials (including hierarchical notation such as collection > series > item), as well as the cost burden of the processes for making information available. Looking at future prospects, I took up the CCR (Connection between Collection and Research) concept. The goal of the CCR is to provide an infrastructure for connecting a variety of specimen collections, not just type specimens (those used as evidence when proposing new biological species), to research results and data, and to create integrated connections between the research cycle and archiving cycle. This will make it possible to link open academic materials information on the Web to scholarly articles and publications or to materials archived in museums, and then make clear their relationships. When making academic materials open, how to bring together diffuse information and link it to actual scholarly activities, namely root materials like specimens, should be a major key. It will further be necessary to create schemes for encouraging and supporting the archiving of physical materials that involve costs. There is a strong need for progress in this field.

Institutional Repositories and DOI: Assignment of DOI in JaLC

Hideaki Takeda
(National Institute of Informatics)

Prior to the digital age, the final research output of scientists meant their scholarly articles, and data was no more than information for writing articles. With the rapid rise of digital archiving, however, an enormous volume of data has come into being, so that today data itself is research output, and articles and data are now becoming integrated. Following on theoretical science, experimental science, and simulation science, we are today seeing the emergence of data-centered science. In the case of simulation science and data-centered science, scholarly articles are no longer the research output but simply serve to introduce the research. The data itself is the research output. If

we ask why research data should be made open, the reasons include (1) sharing research results with society, (2) the public nature of publicly funded research results, (3) the continuity and further development of research results, and (4) ensuring reproducibility.

Among the layers of information infrastructure supporting the distribution of research data, the identifier at the top of metadata is growing in importance. There are many different metadata schema for describing data, and using only metadata to identify data and control its distribution is becoming increasingly difficult. Identifiers today include DOI and also ORCID (researcher identifier) and FundRef (identifier of funding agency), among others. DOI is a service that converts an identifier to a URI where the digital object is found. It was created originally so that publishers could share article identifiers, but today has grown into an identifier of various kinds of digital objects, not just articles. The advantage of the DOI service is that it provides a reliable means of access to content. This is of great importance to all stakeholders, from authors to readers, publishers, and funding agencies.

The service has a three-layer administrative structure, consisting of the IDF (International DOI Foundation) with overall governance responsibility, DOI Registration Agencies (RAs), and DOI issuers. CrossRef, one of whose missions is assignment of DOI to scholarly articles, is also an RA. Another RA is DataCite, which assigns DOI to data sets.

Japan Link Center (JaLC) is likewise an RA. When it was established in 2012, the first phase of its service was mainly assigning DOI to journal articles. Upon moving to a new system in December 2014, JaLC has drawn up policies enabling it to meet various DOI assignment needs in Japan. Of particular note is the expansion enabling DOI assignment to institutional repository content. The direction aimed for by JaLC DOI is to realize a DOI that can cover the entire range of researcher accomplishments. If, for example, DOI can be assigned to all results of the Grants-in-Aid for Scientific Research and other such programs, this should be useful to the researchers themselves and also to research institutions and funding agencies.

A test project to assign DOI to research data was started in JaLC in October 2014. The goals are to set policies on DOI registration of research data and establish an operational flow. It is the first project in Japan linking research data-related organizations across different fields. Research data involves many issues not faced by scholarly articles, such as metadata schema, data granularity, and the relationship between the data life cycle and actors. This project is now moving ahead on solving these issues.

Research output will eventually come to be “data” and the provision of a research data distribution infrastructure will be essential. The DOI is sure to become an important element as an indicator of that distribution.

Panel Discussion

Toward the Expansion of Green Content

Moderator: Hiroshi Horii (Academic Repository Network)

Panel members: Kazuhiro Hayashi (National Institute of Science and Technology Policy) / Yasuyuki Minamiyama (National Institute of Polar Research) / Shunsuke Yamashita (Kyoto University Unit of Synergetic Studies for Space) / Hideaki Takeda (NII)

Ahead of the panel discussion, Kazuhiro Hayashi of the National Institute of Science and Technology Policy gave a brief presentation titled “From Open Access to Open Science: An Overview and Main Policy Issues”; and the moderator, Hiroshi Horii of the Academic Repository Network, introduced initiatives related to academic materials. Summaries of their presentations follow.

HAYASHI: Starting in this decade, global society has moved from open access to open science, as we have entered an era in which ordinary citizens are consciously and unconsciously coming closer to science. Recent years have seen attempts such as in science communication to make a more active commitment to citizens, and a greater variety of stakeholders are becoming involved than were in the open access era.

The 4th Science and Technology Basic Plan does mention the promotion of open access, but the activities in response have been largely limited to creating institutional repositories and supporting digital archiving of academic journals. A major move took place in 2013, when Japan agreed to open up research data. This was at the G8 Science Ministers meeting in parallel with agreement to the G8 Open Data Charter.

Given these trends, what should we do next? Policy measures must not only be made with regard to science and technology but must spread to economic benefits, industry promotion, and educational benefits. Since open access is not

necessarily to the advantage of researchers in all fields, the question of what should be open and what should be closed needs to be considered from the standpoint of national interests. The time may come when those wishing to make data closed will be asked to provide a reason. The library industry will need to pay attention to trends in data journals, and new elements may emerge, such as research data quality and the degree of contribution of data producers.

HORII: At the Academic Repository Network (Re*poN), established in October 2014, people involved in academics in universities, companies, and elsewhere are working on projects to convert materials into data. To date, Re*poN has been creating digital archives of scientific laboratory instrument materials and of educational wall charts used in the Meiji to Showa periods (late 19th and most of the 20th century), and developing the virtual museum project of Kanazawa University. Through surveys and analysis of miscellaneous materials, creation of metadata and digital data, and exchanges across organizations, the aim is to build and maintain the information infrastructure. A case study of research data creation is the genealogical catalog of the Kaga feudal domain. Even though the extraction and digital archiving of data were funded by the Grants-in-Aid for Scientific Research program, the results were only available to the research community Kaga-clan Research Network. The issues for release to the general public include (1) accuracy and completeness as research data for general release, (2) permission for release from the archivers and others involved, and (3) the labor and cost burden of creating public data. Here the importance of the people in charge of forming and overseeing the data can be seen.

In the panel discussion that followed, participants offered their views from various standpoints on the theme, “Expanding Green Content.” Examples are given below for each topic covered.

Significance and purpose of expanding openness to academic resources such as museum materials and research data

TAKEDA: The trend toward openness is greatly facilitated by the changes in the international community. My personal interest is in how libraries and institutional repositories will respond and what kind of meaning they will find. This relates also to the issue of how far libraries can be involved in the research process. Academic fields are likely to become more and more specialized in the future, but I think it will be important to search for areas in common among the fields.

YAMASHITA: Museums have for some time been compiling materials databases and making them available. With the opening of data, this is a good time to think about how to connect the results up to now with other movements.

MINAMIYAMA: Making materials public and organizing them are central to a library’s mission; that is why libraries exist. That role does not change even though the materials are digital media.

HAYASHI: Without the development of industry in academic fields and related fields from openness, there can be no increase in research budgets. Speaking from the standpoint of how libraries should be involved in data, the quality control problem arising with data journals is not a matter of checking the contents of the data itself but relates only to format. Expertise in checking formats is something libraries have accumulated.

The path to openness

MINAMIYAMA: There is an urgent need to draw a line between open and closed, so as to make clear the scope of openness. In the case of libraries, it will be necessary to draw up, in cooperation with researchers, foundational policies regarding, for example, the range of data to be handled in the repositories. As for the practical aspects (and things will go much faster if librarians have a grasp of the main issues), we should first of all take them on in cooperation with graduate students and URA.

HAYASHI: For dealing with data, a data management plan is necessary. The library needs to be involved from the research planning stage, and one suggestion is to provide training for this. For policymaking, cooperation with URA should be carried out as an action plan.

Dealing with orphan museum materials

- What is to be done about materials for which a license cannot be obtained because the creator cannot be contacted?

YAMASHITA: This is a difficult issue. Our current approach is that at the time of call for proposals, we give priority to selecting those that can be made public or for which license processing is possible.

HAYASHI: The proper way to handle such cases is to make use of the Agency for Cultural Affairs compulsory license ruling system; but in the past when similar cases arose with scholarly articles, they were handled by making notification on the Internet. The thinking was that the risk of legal action was small with scholarly articles; but the same may not be true with museum materials, so this approach is not recommended.

FROM THE FLOOR: Since it’s about distributing scholarly information, I don’t see how it could be a problem. Why not just go ahead and make it public, and then if someone complains, deal with it in good faith?

Priorities in making museum materials available

YAMASHITA: I think priority should be given to making available information about type specimens and materials used in scholarly articles.

HORII: One approach to releasing scholarly information would be to start with information giving an overview of the materials, and to release more detailed information in phases based on the situation.

While not all topics could be discussed as fully as desired, the panel members ended the session by making clear their resolve regarding data management.

--From attendees----------

(people affiliated with university libraries)

- This was a highly worthwhile seminar that turned out to be very educational. It was my first time to hear about Re∗poN, and I was grateful to learn so much in just four hours about the latest trends, issues, and news regarding open science and related topics.
- Even though not much was offered that could be reflected immediately in my work, it was stimulating to learn about global trends and what is in store for Japan.
- There was more conceptual material than I had expected. It was useful, though, for getting a grasp of the overall picture.

(university educator)

- I learned about many of the different issues involved in open data.

(university staff)

- It was helpful to learn that moves at the policy level are already under way toward the opening of science data.

(person in university, involved in academic journal editing)

- I learned about world trends concerning the sharing of research data. The seminar made me think about what I can do, as a journal editor, in response to these trends (Publish data journals? Call on authors to provide data to repositories?). My impression, though, is that data sharing is not yet being talked about much among researchers.

(person in corporation involved in university education)

- I listened with special interest to the discussions about moves at the policy level. This kind of overview is necessary when asking how libraries should respond to this major trend.

(person in corporation)

- I learned about the current state of Green Content and future prospects.

(person affiliated with libraries)

- The seminar made me aware of the kind of approach needed to promote institutional repositories, and of the need to think about handling research data.

---------Afterword----------

This is actually my first time to take part in SPARC Japan Seminar, and I'm at the stage of looking around saying, "Hmm, so this is what it's like." As for the topic at hand, we are hearing a lot from all sides about research data, but I came away with the strong feeling that the more you think about it, the more you realize the differences in the handling of data and scholarly articles. At the same time I feel more and more that this is a hugely important matter. One takeaway is that I am no longer sure articles will continue forever to be central to scholarly information.

Taro Misumi
(Chiba University Libraries)

We started planning this seminar on the theme of open data, but while we were making preparations there were many rapid developments among related institutions in Japan, which kept us busy day after day trying to keep up with them. I would be most pleased if we succeeded in sharing with everyone the joy of talking about future data management from a variety of standpoints.

Yasuyuki Minamiyama
(National Institute of Polar Research)

Though it was my first time not only to serve in the planning WG but to attend a SPARC Japan seminar, it was a stimulating experience in many ways. Since in this seminar the scope of openness was expanded to museum materials (!?), I wanted to work with Mr. Yamashita of Kyoto University, also a member of the Academic Repository Network, to inform as many people as possible about the current situation of museum collections and the challenges they face. For analog materials such as ancient manuscripts and specimens to be made available over the network, the efforts and cooperation of many people are necessary. I want to apologize for not doing a better job of moderating the panel discussion so that it held together more, but it is my sincere hope that there will be more opportunities in the future to bring together people from a variety of standpoints for frank and enjoyable discussions on open access to academic information.

Hiroshi Horii
(Academic Repository Network)