Metadata expert from Japan

Juha Hakala
Senior adviser
The National Library of Finland
juha.hakala@helsinki.fi
https://orcid.org/0000-0003-1067-5020
ISNI: 0000 0004 1662 5064

Introduction

Metadata has not been a popular research topic in Finland, and it is hardly main stream even abroad. However, there are library schools in e.g. the U.S.A. which analyze the production and usage of metadata. In Japan, professor Shigeo Sugimoto has done metadata research since 1990's.

Professor Sugimoto spent a week in Finland in November 2018, visiting the National Library, CSC – IT Center for Science and the Institution of Information studies and interactive media in Tampere university. I had a great pleasure to interview an old friend – I met Shigeo for the first time in a Dublin Core conference in 1997 in Helsinki.

Shigeo Sugimoto became an associate professor in the University of Library and Information Science in 1986, only a year after he got a PhD in engineering. Since 2002 he has been a professor in the Faculty of Library, Information

and Media Science in the University of Tsukuba. He studied computer science, but became an information scientist already in the beginning of his academic career. This shift was a little bit of a culture shock, not least because most computer scientists are also in Japan men, whereas the majority of the students at library and information science were women.

In addition to his successful academic career Shigeo has been entrusted with many national and international positions of trust. Some of them have been listed in the end of this article. These responsibilities have also had an impact on his research interests.

The interview is followed by some thoughts based on professor Sugimoto's answers.

Interview

Q: Dear Shigeo, due to retirement, this trip to Finland is the last one you will make as a university professor. Why did you decide to come to Finland?

A: I've been willing to visit Finland for many years since my experience in DC workshop in 1997 was so impressive and I knew your activity in digital archiving and preservation.

Another reason is to visit and see iSchool in Finland (University of Tampere) as a long-time participant of information schools collaboration activities.

Q: In Finland, it is possible to study library and information science both in universities and in polytechnics. What about Japan? How many information schools and information science students are there?

A: There are some departments/schools of LIS, for instance U. Tsukuba, Kyushu U., Keio, and a few more.

Librarianship certificate programs for public libraries are provided by more than 200 universities and junior colleges. These programs are taught by two or three full time faculties plus part-time instructors (and adjunct faculties). All such programs have to be endorsed by Ministry of Education.

Local governments are suggested by the national government to hire to the public libraries under their responsibility fulltime employees who have the librarianship certificate.

Q: How many staff members and (doctoral) students does the Tsukuba faculty have?

A: There are 60 full time faculty members, of whom four are non-Japanese faculties. There are also a few Japanese faculty members who got their PhD's abroad.

PhD and master programs are three and two years long, but many PhD students need more time to complete their studies. There are about 100 master students and 50 PhD students in all; of them, about 5-10 of the former and 3-5 PhD of the latter study in my lab.

Q: What are the main research interests of your information school?

A: There are many, because we have 60 faculties. Roughly, some major topics:

Conventional LiS: Library Management, Social Policy, Bibliometrics, Scholarly Communication

Crossover LIS & IT: Metadata on Web, (Lib) Info Systems, Digital Libraries/ Archives, Information Seeking/IR

More IT oriented: HCI, Data Engineering, Social Network Systems, Data Analytics, etc.

Other areas: Communication, Education, Content Technologies (New Media-Technologies and their application) and Legal Issues.

Q: You have been actively involved with metadata models and schemas. How did that happen?

A: Through Digital Library Initiatives in 1990s and Dublin Core Metadata Initiative.

Q: University of Tsukuba has been one of the key partners in building the information system which brings together diverse information about the 2011 earthquake which devastated Fukushima nuclear power plant and several coastal communities. How did the researchers get involved, and what was your contribution?

A: National Diet Library (Parliament library) has been the main hub to collect and archive digital records (text documents, still images, etc.).

We (Tsukuba) have been doing research but we have not been responsible of establishing the actual services.

A colleague professor of mine, Prof. Tetsuya Shirai who was an archivist, is collaborating with Futaba Town, Fukushima prefecture to archive their materials collected while the whole town had to be evacuated for some years after the nuclear disaster.

Q: How does this project compare with other initiatives with which you have worked, such as a metadata model for manga?

A: As a project, there is no direct connection but what I have learned from these projects is that metadata aggregation is a key technology to bridge the gap between end-users and providers (e.g.,memory institutions). End users prefer data of higher level / larger granule such as Wikipedia article but metadata provided by memory institution is primarily manifestation- or item-based. Bridging the gap between these ends would be required. Another issue is to link metadata (=resources) across temporal gaps, e.g., connecting records made in 1950s and those made in 2000s.

Q: What role does metadata have in the Internet, which relies on Google for information search?

A: Snippets provided by search engines such as Google are the metadata which the users read first, since Web information retrieval begins usually with these services.

Wikipedia articles and/or end-user oriented webpages describing target resources would be the second step, since search engines give them a high priority.

Institutional metadata, e.g., library catalog and museum catalog, would be the last hop to access the materials stored in (virtual) libraries/ museums. Users may find links to these resources from e.g. Wikipedia, which has links to various library services. But such links are not common enough.

I think bridging the second and the third step is crucial for the metadata community. However, there are many technological challenges there. Library metadata concentrates on manifestations and items, and there is a big gap between Google snippet on some topic, and library metadata describing a book about that topic. This gap cannot be easily covered by technical means, we need metadata which "explains" the gap between a snip-

pet and metadata record, which has been removed from its original context in a library database.

Q: Metadata models and schemas such as Dublin Core are international. What kind of impact has this had on your work?

A: There are several different aspects/viewpoints on this.

> *1. I think that people at memory institutions (professionals) tend to see metadata as basically manifestation- or item-based, even if they know about FRBR. What I learned from my metadata research is that granule of Items depends on how people recognize their objectives to find, access and use information. In general granule size at libraries and museums is very small.*

> *2. DCMES has been accepted by these professionals (who used to live in their silos) as a standard and they understood DCMES is crucial for interoperability across communities. This is OK but I think that there is a gap between those professionals and end-users, who are not very familiar with manifestation- and item-based metadata in, e.g., library and museum catalogs.*

> *3. Most metadata standards define schemes to describe their target objects, which may be either hand-held or electronic (immaterial). Combining those different schemes to achieve semantic interoperability is required on the net, but that is still not an easy task. Basic process models to combine different types of metadata are required.*

Q: You have been one of the key figures in the Dublin Core metadata initiative for more than 20 years. How has this influenced your thinking about information systems and metadata?

A: I'm not sure whether I was a key "figure" for DCMI...
*   Metadata is basically description of something. But, I learned from DCMI activities that combining metadata across somethings (communities, time, contexts, convention, viewpoints, etc) is a key issue. "Librarians are manifestation-centric but end-users on the net are not. How can we bridge the gap between libraries and end-users?"*

Q: Digital preservation is one of the key challenges information society is

facing. How should we get prepared? Do you think Japan has done enough?

A: Japan is far behind from what is required. The current long-term management of digital resources can be described as optimistic.

In order to improve the situation, I think the first priority is to get a shared understanding about:

- *Perfect long-term preservation of either digital or physical resources is not possible*
- *Preservation is a maintenance activity over the years and generations*
- *There is a need for consensus on preservation policy by institutions and governments and with general public*
- *There is a need for broad understanding that digital preservation has many different aspects and that it needs different functions/services*

Q: During this visit, you have also seen what Finland is doing in this area. Do you think we are on the right track?

A: Yes. In particular, the roles of CSC and their functions were very impressive.

Q: What advice would you want to give to information scientists in Finland?

A: This question is hard to answer because I know only a few information scientists in Finland.

Social environment of Finland is quite different from Japan admitting the diversity in Europe. However, I think that we (information scientists in Japan and Finland as well as those in other countries) need to collaborate with people in other regions in the global information environments to bridge gaps across culture, languages and time.

Q: Thank you – Arigatou gozaimashita!

The role of metadata in the information society

Neil Gaiman has said: "Google can bring you back 100.000 answers. A librarian can bring you back the right one". The librarian cannot accomplish

such a feat without understanding the user's requirements and the information resources available. Descriptive metadata does not help with the former, but it plays a crucial role with the latter.

Even traditional manifestation-oriented metadata such as MARC records describing books make it easier to find relevant resources, but this metadata should be complemented by metadata about collections. The problem with such metadata has been that it is difficult to integrate it with manifestation-oriented metadata and IR systems which utilize such metadata.

Project Kokoelmakartta (Collection map, `http://www.varastokirjasto.fi/kokoelmakartta/`) was a pioneering collection description initiative in Finland. The project built a set of tools (including DC Collections Application Profile -based metadata format, cataloguing template and guidelines) which are still available, but the DSpace-based database of collection descriptions the project created was not maintained after the project and later it was lost.

Using Shigeo Sugimoto's terminology, increasing granule size is important, but it may also be a challenge. Instead of making the granules bigger, libraries have been making them even smaller – in Finland, a lot of resources have been spent on cataloguing e.g. journal articles. This is definitely useful, but does not enable the users (or librarians themselves) to understand the collections libraries have. Such information may be undocumented – silent knowledge of curators and other collection specialists – or "hidden" in printed guides and web pages. We do not have efficient means via which to make collection metadata available in search systems such as Finna (`https://www.finna.fi/`).

National libraries and archives are among organizations preserving electronic resources for future generations. Traditional descriptive metadata and current cataloguing rules do not support this. Libraries must create administrative metadata: technical metadata describing the properties of files, and preservation metadata describing actions such as file copying and migration carried out in order to preserve access to the intellectual content.

Granule size is again a challenge here: instead of cataloguing just individual files, it is necessary to describe file formats. Digital preservation community has built services like PRONOM (`http://www.nationalarchives.gov.uk/PRONOM/Default.aspx`) to meet this requirement. So far these services, or digital preservation systems, have not been integrated to library systems in a satisfactory manner. Such integration would enable a librarian to see for instance which applications could or should be used to render a digital resource she is cataloguing, or to choose preservation actions required before the resource is added to library collections. If the file format is at risk, it may be necessary to either migrate the resource or specify an emulator with which the resource may be used in the future.

Due to digitalization the role of metadata has changed in information retrieval. Full text indexing makes it easy to locate resources such as newspaper articles which used to be difficult to find it the past. But the need for creating and investigating metadata has not disappeared. On the contrary, we need metadata more than ever to preserve our cultural heritage to future generations, and to enable the users to find resources which are the most relevant for them. It is to be hoped that the information scientists will continue to help the libraries in the development of systems and services which will meet efficiently the users' requirements.

Appendix

Occupation: Professor, Faculty of Library, Information and Media Science, University of Tsukuba

Education/Degrees:
Bachelor of Engineering 1977.3
Master of Engineering 1979.3
PhD in Engineering 1985.7
All degrees from Dept. of Information Science, Faculty of Engineering, Kyoto University, Japan. Information science equals computer science in Japan.

Jobs:
1982.6 – 1983.9: Research Associate, Dept. of Information Science, Faculty of Engineering, Kyoto University
1983.10 – 1986.9: Research Associate, University of Library and Information Science (ULIS)
1986.10 – 1999.2: Associate Professor, University of Library and Information Science
1999.3 – 2002.9: Professor, University of Library and Information Science
2002.10 – present: Professor, Faculty of Library, Information and Media Science University of Tsukuba

Scholarly and Professional Activities:
4. International Conference Committees (Steering Committees)
   - Steering Committee member of Joint Conference on Digital Libraries (JCDL)
   - Steering Committee member of International Conference on Asia-Pacific Digital Libraries (ICADL)
   - Steering Committee member of International Conference on Digital

       Preservation (iPres)

5. International Organizations
   - Member of Governing Board (formerly, Oversight Committee) and Advisory Board, Dublin Core Metadata Initiative
   - Chair, iSchools consortium Asia-Pacific Chapter, 2016.7 – 2018.3