

(2014年11月7日開催の2014年度国際図書館協力シンポジウムの音声を文章化した。#数字は和訳と対応)

From Wooden Chests to Digital Files: The Changing Face of Scholarly Collections at Cambridge University Library

Anne Jarvis (University Librarian, Cambridge University Library)

Thank you, and good afternoon. As University Librarian at the University of Cambridge, I am pleased to have the opportunity to share my thoughts and experience in relation to the changing face of scholarly communications at Cambridge University Library.

I would like to begin by introducing you to the University of Cambridge and then focus on the library itself, its history, and how it is positioning itself in an ever-changing environment with particular reference to the Cambridge Digital Library. I will then conclude with some thoughts about the library of the future.

[#: indicates slide number]

#2

The University of Cambridge is one of the world's leading universities and one of its oldest. It celebrated its 800th anniversary in 2009, marking the legacy of eight centuries and it is determined to remain among the world's greatest universities. Its reputation for outstanding academic achievement is known worldwide and reflects the intellectual achievement of its students as well as the world-class original research carried out by the staff of the university and the colleges.

#3

The modern university is a confederation of schools, faculties, departments, and colleges. There are six schools, which each form an administrative grouping of faculties and other institutions. There are arts and humanities, biological sciences, clinical medicine, humanities and social sciences, physical sciences, and technology. There are 31 colleges where staff and students of all disciplines are brought together. Students live, eat, and socialize in one of the colleges. Along with class-based lectures, the 11,000 undergraduates receive college supervisions which are small group teaching sessions which are regarded as one of the best teaching models in the world.

#4

Cambridge University Library is formally described as a non-school institution. It is part of a tripartite system in Cambridge consisting of 95 separate libraries. There are 31 college libraries,

18 libraries that are part of the University Library, and a further 44 departmental and center libraries, many of which are now joining the University Library. Across the libraries, there is a long tradition of collaborative working which helps libraries to avoid duplication of effort and resource.

#5

Cambridge University Library's history spans six centuries. Manuscripts were originally kept in chests such as the picture here until in the 1420s, a place was created for the more than 100 volumes which had been held. Over the time, the collection grew with many additions in the 17th and 18th centuries, and by the 19th century, new buildings were required to contain the library. Two of the key events in the library's development occurred in the early 18th century.

The Copyright Act of 1710 established the library's copyright status. It gave publishers copyright protection on certain conditions, one being they had to send copies of their books to a number of privileged libraries, Cambridge being among them. Then in 1715, the Library received by gift from King George I, the magnificent library collected by John Moore, Bishop of Ely who had died the previous year. The library of John Moore, which from the circumstances of its donation became known as the Royal Library of the University Library, was renowned throughout Europe and contained some 30,000 volumes of which 1790 were manuscripts. This donation brought to the library a comprehensive collection of books of all periods. There were 470 incunabula in all departments of literature and learning including many of the library's most valuable items.

#6

The Library continued to grow and by the 1920s, it was noted that the University's old library building was full with its contents at that time occupying 20 miles of shelves. More than a decade of talks and fund-raising led to the library building of today with its iconic 157-foot tall tower designed by the renowned Sir Giles Gilbert Scott who also designed the United Kingdom's telephone box. It is like a telephone box.

Today, Cambridge University Library takes its role as a legal deposit library very seriously indeed. This was previously called a copyright library. It forms part of our national archive. Many libraries regard the printed text as a replaceable item. Cambridge University Library keeps only one copy of each edition and aims to preserve it forever. Uniquely, among the deposit libraries in the United Kingdom and Ireland, Oxford, Cambridge, the National Libraries of Scotland and Wales, Trinity College Dublin, and the British Library, Cambridge stores two million of its books on open-access shelves allowing the readers the facility of browsing among works on related subjects. It is therefore one of the largest open-access libraries in the world.

#7

The Copyright Act was replaced by the Legal Deposit Act in 2003 which brought electronic publications and other print publications into the scope of the previous legislation. However, in spite of all the predictions of the death of the book, the increasing electronic availability of resources has not yet been matched by any significant decrease in traditional monograph publishing though there has been a significant change in scholarly journal publishing. Each year, nearly two miles (or 3.5 kilometers) of extra shelving has to be provided for the 100,000 books received by the Library. This puts immense pressure on our restricted resources, but the legal deposit collection, which represents about two-thirds of our annual intake, is one of the Library's greatest strengths. The regulations related to Legal Deposit have recently come into force and this will mean a new range of services for the Library, as it can only receive one item in one format under these regulations. So, when the journal is received in digital format, the print equivalent has ceased to be deposited from that date. The same applies to books.

As noted earlier, the library continues to grow and now holds approximately eight million items, occupying 100 miles or 160 kilometers of shelves all in one building. In the 21st century, the Library's digital growth is equally prodigious. Future librarians will be as concerned with terabytes of virtual space as with miles or meters of physical stock. Browsing this material, whether in person or online, has been described as walking around the world mind. Over the course of the past six centuries, the University Library's collections have grown from a few dozen volumes on a handful of subjects into this extraordinary accumulation of several million books, maps, manuscripts, and journals, augmented by an ever-increasing range of electronic resources. They cover every conceivable aspect of human endeavor across 3000 years and in over 2000 languages. From its beginnings, as an asset for a tiny community of theologians and canon lawyers in the medieval university, the library's mission has expanded to serve the international scholarly community and now, through its digitization projects, to reach new audiences across the world. Great collections count for little if they cannot be discovered and explored. So, while we conserve this unique cultural heritage for the future, we are simultaneously finding new ways to share it with the present generation by building a digital library.

#8

Until recently, we knew exactly what the Library was for, who we served, the scholarly community and how we should deliver the library service. The model of a research library was very much that of a stand-alone service provided to the university. However, the rapid development of technology means that we are now dealing with enormous change as we seek to deliver an exemplar library service to a worldwide virtual audience whilst maintaining excellence in all we do. Those who predict the demise of libraries in the face of emerging technology fail to understand how the libraries such as ours are rapidly evolving to lead and support both interactive learning and

knowledge creation across the physical and digital realms. It has been estimated that 89% of students are now using smartphones in addition to laptops and tablets. Therefore, libraries need to offer learning opportunities across multiple mediums that allow students to access local and global networks of information. That access allows not only access to knowledge, but also the ability to easily create and share knowledge. This is where the 21st century teaching, learning, and the new creation of knowledge intersect. Libraries must now foster a positive ecology of relationships, connectivity settings, and tools layered together to foster discovery and learning within the context of a dynamic academic framework.

New forms of scholarship, research, and teaching present significant challenges to libraries today and in planning for the future. Readers want access to a wide variety of formats delivered to a range of devices at their own convenience, anytime, anywhere. In an increasing open world, they want to reuse, mix, annotate, and share information. The boundaries between formats and types of content (articles, monographs, and data) are blurring. Authors want to disseminate ideas more quickly and more widely to achieve impact for their work. Research funders are increasingly focusing on the dissemination of the research they fund. Our universities are looking to new ways of showcasing the reach and breadth of the intellectual outputs produced by their researchers and students. Reflecting this, Cambridge University Library is advancing on three fronts; the digital, the analogue, and developing services and partnerships using our skills and resources. While forging new relationships and ways of working together, we aim to provide our users with innovative and responsive information services.

#9

The Library therefore now extends beyond its shelves as it serves up electronic resources based almost anywhere. Through its repository DSpace@cambridge and the Cambridge Digital Library, we are making our unique collections and born digital collections available to all. DSpace is the University's repository for research publications and data capturing, storing, disseminating, and preserving digital materials created in any part of the University, while the Cambridge Digital Library enables scholars to work interactively with collections, not only accessing them without ever having to travel to Cambridge, but annotating and transcribing materials, feeding back research, and sharing their findings with virtual communities across the globe.

Today, alive to the opportunities of digital technologies, the Library is transforming scholarship by opening up access to many of its remarkable collections through digitization. Far from undermining the importance of our collections, the digital revolution has presented an opportunity for the Library to expand its reach by enabling us to open up our collections to new audiences and new methods of study and analysis.

The dynamics has thus changed from building collections and requiring users to come to

the physical library to pushing our content out and seeing it used by new audiences and in new ways. The library is no longer collection-driven, but use-driven, and this demands that our content becomes visible and usable in the digital world and with appropriate tools and licenses to support this.

#10

The most visible part of our digital library strategy to date has been the Cambridge Digital Library containing 25,000 items with a 250,000 pages of content. It receives an average of about 50,000 unique visitors to its site each month and the number of new users and returning users are growing as more content is added and the digital library becomes increasingly embedded in the Internet.

We launched the Digital Library in late 2010, but it had a long prehistory. The Library has been commissioning and creating photographic representations of items in the collection for more than 100 years, and it had large microfilming programs for much of the second half of the 20th century. Our Imaging Services department embraced digital technology about 20 years ago and has been completely digital for the past 10 years.

Our initial digitization strategy stated that, through digitization, the Library aims to create, manage, deliver and preserve high quality digitization content that addresses the needs of those undertaking research and education, locally, nationally, and internationally. This is still our goal.

#11

We identified the main aims or benefits in undertaking a digitization strategy as: improving access to items or collections, particularly for off-site users or where access to the material is restricted: improving discovery of the Library's collections; supporting preservation of fragile items by minimizing the need for physical handling; enhancing our content's research potential by offering increased functionality or utility such as full-text searching and improved legibility; enhancing our content's teaching potential by providing materials in forms well suited to teaching and learning activities; enabling aggregation of content, bringing together related materials from different collections or institutions; generating income to recover costs and support further digitization activity; developing skill and capacity within the Library so that it is well equipped to respond to further opportunities for digitization; and finally, raising the library's profile as a center of expertise in digitization, within the university and beyond.

I believe we have begun to achieve all of these aims. And there is a further benefit of engaging digitization and that is developing partnerships. Partnerships have been critical to the Library achieving everything we have done within the program and we see it as essential to the future.

Cambridge has such rich and vast collections that a key challenge for us as we embarked

on our digitization program was to know where to begin and what material to focus on. We were determined very early on that we should be concentrating on our unique materials or very rare materials which would be unlikely to be digitized elsewhere. Other Libraries and organizations like Google, Microsoft, and the Internet Archive have invested considerable resources in the digitization of published items and so it made sense that we should concentrate on our special collections.

An important factor in determining our priorities for digitization was whether the material was in demand for research or teaching. Here, we relied on input from our academic colleagues. However, we also had to identify material that was not currently being used, but had great research or teaching potential. Many of our valuable special collections are not well used because they are not catalogued or very visible.

Copyright was also another important consideration for us, since we were targeting manuscript material. Under the UK copyright law unpublished works have a long duration. The availability of funding was naturally a key determinant. Although we are a great library, we do not have large reserves to pay for or subsidize our digitization. As with every other research library in the United Kingdom, we needed to find external resources to take forward our program. So, it is very important that we should seek to match our priorities against those of the funders. We needed to be proactive in seeking funders and agile at responding to opportunities as they arose.

There have been three main sources of funding for our program over the past years; direct government funding, philanthropic donations, and research council funding, but we are also seeking to build up other sources including the commercialization of some of our content in educational aggregations and the sale of images and licenses to publishers. This mixed funding model has been very important during the recent period of economic recession. Many other United Kingdom libraries have become dependent on direct funding for digitization provided by such bodies as the Joint Information Systems Committee in the UK and the National UK lottery. As this funding has dried up, activities have had to be scaled back. Cambridge University Library has been able to fund digitization through donor funding and particularly, through partnerships with researchers at Cambridge and beyond which has enabled us to access significant research funding for digitization.

We regard this alignment with researchers as highly beneficial not only in funding cataloging and digitization of priority collections, but enhancing them by linking in research outputs, such as scholarly editions of works from our collections. It has also enabled us to provide better support for our academic colleagues helping us to sustain some of the research outputs. As a result, we have become an active and valued participant in the Digital Humanities Network at Cambridge.

Another consequence of our collaboration with research colleagues has been that we have had to provide a richer online platform than the simpler commercial or open source digital library platforms that were available in 2010 and indeed in 2014. We have hired computer programmers and invested in a development program. This is an ongoing commitment we will need to make in order

to ensure that we are providing the tools and experience that our users require.

But here too we have tried to proceed strategically and in partnerships with others. Wherever possible we have built our infrastructure from existing tools. Our current platform draws heavily on work undertaken by the California Digital Library and in particular its Extensible Text Framework (XTF) toolset. This provides our search and translates metadata and transcription data into formats for indexing and display. It enables us to bring in data in a wide range of library archive and Digital Humanities formats and use them within the digital library.

We have also been working with other major UK libraries developing their own platforms, particularly the Bodleian at Oxford and the Wellcome Trust Library sharing knowledge and technologies. And we are currently working with a wider, international pool of research libraries to develop support for International Image Interoperability Framework (IIIF) which should make it easier for digital libraries to share and exchange images, metadata, and transcriptions or annotations related to content. This is an important development for us, because in addition to providing content through our own online platform and websites, we want others to be able to take our content and use it in their own projects.

#12

This research collaboration and sharing of content can be exemplified in the projects we have undertaken of one of our greatest scientific collections, the archives of Isaac Newton. Our first major project for the digital library was the Newton Papers. We secured significant funding for the development of our infrastructure and digitization and wanted to launch the program with a high profile collection to make a mark and attract further funding. Newton's papers were an obvious choice, but in addition to digitizing it, we approached a long-term academic project based at the University of Sussex in the UK who were producing a scholarly edition of Newton's works to see if we might be able to link our facsimiles with their transcriptions.

#13

The Newton Project developed mechanisms to enable us to automatically harvest their transcriptions and we developed tools to enable them to take our images from us and to build links between the two online presentations. We saw each presentation as complementary with Cambridge University Library providing an archival view for Newton's papers and the University of Sussex providing a textual approach to Newton's works. We also saw the arrangement as beneficial in enabling people to discover both websites and in providing long-term sustainable online delivery for the outputs of Newton project if the University of Sussex's ongoing funding became uncertain, which is actually currently the case.

Now five years on, Cambridge University Library is working in a very similar way with an

external project producing a scholarly edition of Darwin's manuscripts, another of our great holdings. The Darwin Manuscripts Project is actually based at the American Museum of Natural History. I will be talking a little bit more about Darwin tomorrow. This time, the project is being funded by the American National Endowment for the Humanities, the major US research funder for the humanities. Again, we are automatically harvesting transcriptions and displaying these alongside facsimiles and we are providing images of the manuscripts for use by the project. In addition, we are bringing in a third project, the Library's own editorial project based around Darwin's correspondence. Over the next few years, the Cambridge Digital Library will become the key hub for Darwin's archive and associated scholarly activity taking on a coordinating and linking role and offering long-term sustainability for these efforts.

Another good example of how we are building the digital library through collaboration and strategic partnerships is the Board of Longitude Project which ran from 2011 to 2013.

#14

The Board of Longitude Archive is an important collection related to British science and technology and exploration in the 18th and 19th centuries as the scientific establishment sought to determine better ways of discovering a ship's location at sea. The UK Arts and Humanities Research Council, the main funder of humanities research in the UK, had given major funding to Cambridge academics and to the National Maritime Museum in London to research the history of the Board and its activities. The Library then proposed a digitization project to run alongside the research project digitizing all the relevant documentation at Cambridge and at the National Maritime Museum and putting it online to support the project and to open it up to other researchers to use.

#15

This was a very exciting and very rich collaboration. The researchers helped inform the cataloging and they wrote the scholarly introductions and articles to help contextualize the digital archive. Once digitized, the research teams were more easily able to search the archive and this had led them to new discoveries. The collaboration with the museum was also very important for us. We brought together all the digitized archives on the Cambridge Digital Library, but we also linked it to the museum's collection management system connecting the documents and the associated artifacts.

#16

We made use and learned from the Museum's public engagement team in promoting the digital collection to schools and the general public. There is much that libraries can learn from museums in the UK in promoting collections and reaching wider audiences.

#17

One of our digitization projects is not just aligned with research activity, but is a research project in itself. We are using sophisticated text mining and natural language processing to provide enhanced access to our important Genizah collection. The Cairo Genizah is a remarkable collection of manuscripts and manuscript fragments from a Jewish community in North Egypt. We have nearly 200,000 dating from the sixth to the 19th centuries. This had already been digitized before the digital library program as part of an external project, but we have begun publishing them with metadata on the digital library platform.

We estimated that it would take many years to provide metadata for this collection and came up with the idea of using data mining techniques to analyze the extensive literature around the collection, enabling us to automatically generate metadata for the collection and provide automatic recommendations for scholars exploring the collection. The project was funded by the Mellon Foundation in the US and had just completed phase I using the data mining to generate the metadata. We are about to embark on phase II developing a more sophisticated search and display based on this data. This should be available in early 2015 and then we will seek to use this technology with other collections within the digital library.

#18

We are continuing to add collections and the latest is actually a selection of manuscript and early printed works from our Japanese collection. This was funded through a generous donation from Professor Ishii of Kanagawa University.

#19

Cambridge holds one of the five major Japanese collections within the United Kingdom. We hold about 90,000 volumes of Japanese books including around 10,000 early Japanese books. There were two main phases to the formation of our collection, important early works were acquired in the 1910s and then, in the late 1940s, the Library started to acquire modern books. Now the online collection represents the beginnings of third phase, the digital collection. Although currently modest in size, this is something we want to develop in partnership over the coming years.

#20

While our digital library program has achieved much success, there have been a number of challenges along the way and there remain a number of challenges for us. I have highlighted partnerships as a key component of our success, but I should acknowledge that this can also be an area of challenge. Partnerships take a long time to cultivate and need to be carefully maintained. Partners will often have different priorities and these need to be clearly identified and balanced.

Formal agreements are necessary, but the informal and personal relationships are often crucial to the success of collaboration.

One of our biggest challenges is sustainability. We have a fast growing store of content and an increasingly sophisticated technical platform. How do we preserve this in the face of continual technical change and growing expectations on the part of our users? How do we maintain sufficient skills to take the program forward? And how do we fund it? Although we have been quite successful in building the program through our mix of funding sources, we know we need to devote more of our core funding to this activity. What are the appropriate business models we should be employing?

Digital preservation is also a challenge. As our collections of images, metadata, and other content grow, they will need careful management. At present, digital curation can pose more challenges and uncertainty than the curation of our physical collections. Another challenge we face is to meet the needs of the multiple audiences we are now serving. In making our content available digitally, we are supporting scholars who we know will benefit from easier access or easier search. But our content is now reaching much wider audiences who have different needs and interests. How can we support all of these diverse users and needs? Should we be putting our limited resources into making more content available, or providing greater contextualization and interpretation to the resources we have already?

Finally, there is a challenge here in the role of the Library. The Cambridge Digital Library is a broad platform which goes well beyond the catalog providing online facsimiles and in many cases scholarly representation of the archives or editions of works. The Library therefore is taking on new roles as research partner and also publisher. So, how do we equip and re-skill ourselves to perform these roles? Digitization has been a key driver for developing our partnership agenda. Will forging more international collaborations help forge new scholarly relationships as well?

In this presentation, I have demonstrated that Cambridge University Library is embracing opportunities provided by digital technologies. Today, the world of higher education and the accompanying digital information environment is one in which experimenting and innovation will continue to be constant though the outcomes may not always be so predictable. It is therefore critical that in such an environment libraries re-imagine the nature and scope of the services provided, work in partnerships so as not to reinvent the wheel as we would say, duplicate efforts, and embrace the possibilities inherent in new technologies.

As the Library moves more into the networked environment, the question here is whether or not our reconfiguration will be dictated by changes in academic scholarly print collection? That is after all what we have built our reputation upon.

The British Library's vision is that by 2020, 75% of all titles worldwide will be published in digital form only or in both digital and print. As digital formats become the norm, I believe that our rich resource of physical content will become even more precious. It is thus vital that we

continue to develop our world-class stewardship skills in conservation and preservation for both print and digital collections.

#21

To date, the Library's reputation has been concentrated upon the depth and breadth of our collections. While I believe that the power of the original will continue, in the future, the Library will fulfill an ever-wider range of activity to support the research lifecycle and the student experience. We will develop more partnerships with the academic community, partnerships with other libraries of similar mission, both nationally and internationally, and commercial partnerships. As with many libraries, Cambridge University Library will experience continued selective collecting, but in a wider range of formats, a move from collecting to connecting ownership no longer the dominant feature, manage the transformation to a digital environment. This is an organizational and managerial challenge. It is not a technical issue. Continue our increased outward focus. No longer think about control, think about how we enable discovery, and ensure continued access through preservation.

I believe that organizational agility is key to success. The speed of change is not going to slow down. Yet, while things are changing, our values, service provision will remain the same. Academic libraries will survive as long as there are universities. However, libraries cannot thrive without aligning their workings directly to the core mission of their host institutions. Digital technologies provide an opportunity for libraries as we reallocate our resources and services from basic information and directional assistance to collaborative and in-depth research partnership. We therefore need the culture of our libraries to be ever open to change and innovation and our staff must continue to be willing to embrace new skills and competences.

#22

Cambridge University Library has recently celebrated 80 years in our current library building and I would like to close this presentation with words used by King George V at its opening in October 1934. I think they capture something of the challenge we face as libraries in the 21st century, continuing to do what we have always done well, but also embracing the future in a dynamic way. "A Library such as this is both a powerhouse and a testing station of educational activities. It inspires and helps the work of specialists in every field of research and it provides from generation to generation, the standards of truth and sound learning by which the fresh products of science and scholarship are tried. It is a workshop of new knowledge and a storehouse of seasoned wisdom." Thank you.

(Moderator) Thank you very much, Ms. Jarvis.

The next speaker is from the Zhejiang University Library and CADAL project by Dr. Huang Chen, the Vice-Librarian from the Zhejiang University Libraries.