Facing the future of the Research Library: managing and sharing data in China

🕐 December 4, 2018 (http://blog.ukdataservice.ac.uk/managing-and-sharing-data-in-china/)

👤 Neil Dymond-Green (http://blog.ukdataservice.ac.uk/author/neil-dymond-greenmimas-ac-uk/)

💬 Leave a comment (http://blog.ukdataservice.ac.uk/managing-and-sharing-data-in-china/#respond)

Louise Corti reports on her recent time spent at Fudan University, Shanghai as part of their centennial celebrations, which also looked forward to the future of the research library and research data management.

This autumn I was honoured to be invited to give a keynote talk at two events held by Fudan University Library in Shanghai, renowned for being a forward-thinking research library with a great reputation throughout the world.

Fudan and Shanghai

Shanghai is a massive buzzing city, with a population of around twenty-four million, a thriving and business district, and seven ranked universities, with Fudan appearing in the top three.

In Shanghai I was very fortunate to experience the friendliness of all those I encountered. Trying to walk off my seven hour jetlag, I strolled for miles on the first day I arrived, navigating the streets and subway partly by map and partly with the help of young people, who were always more than happy to help in doing basic things such as buying a metro ticket, locating a subway entrance or exit; showing my current location on a map. While road and metro stop names are in English, little else is, which makes getting around without a Chinese-speaking guide a challenge.

I strolled along the two-mile pedestrian walkway along the Huángpǔ River with a fabulous view of the central business district, and its looming towers. A trip up the 636m tower with the fastest elevator in the world let me appreciate the geographical enormity of the city and its densely populated residential tower blocks.
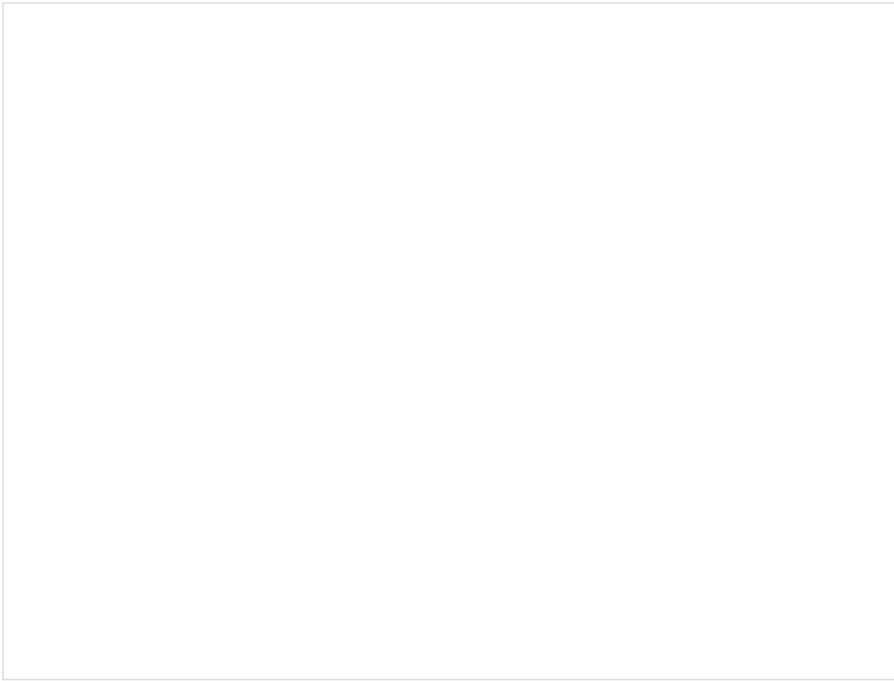
Shanghai

All photos in this post: Louise Corti

Fudan University Library Centennial Celebration: Facing the Future of the Research Library

To mark one hundred years of its prestigious Library, Fudan University invited more than three hundred delegates from libraries across China and those beyond who host Chinese archival collections, or who have donated collections to Fudan.
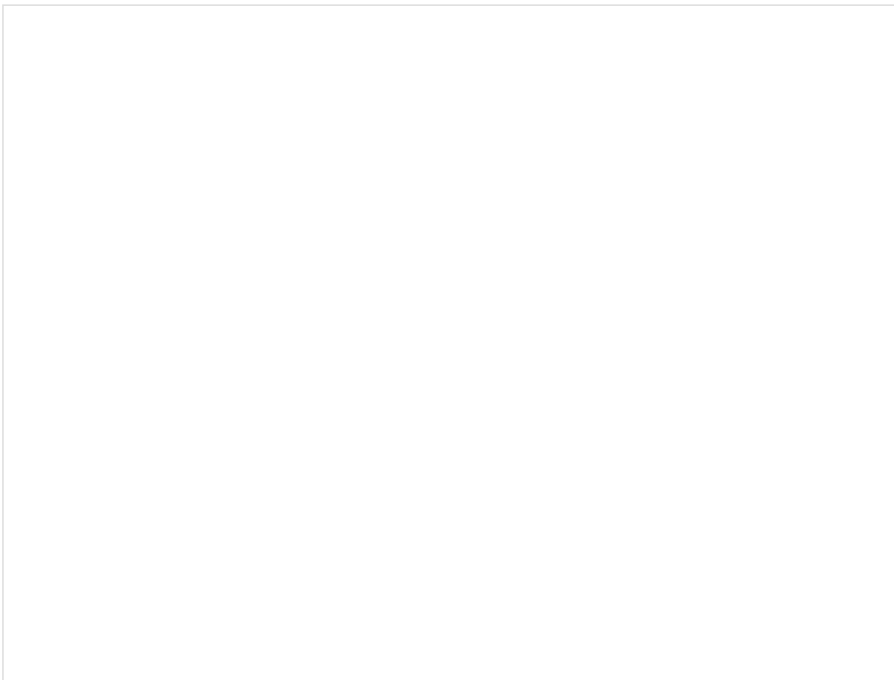

Introduction slide

Jiao Yang, Party Secretary of Fudan University, welcomed delegates and spoke of the success of the past hundred years which has paved the way for restoring historical knowledge and cultural heritage for the next centenary.

She reflected that the library has become one of the top priorities for the University, aiming to build on their already-rich heritage to cultivate a professional team of librarians in preparation for the future. The Ministry of Education's agenda drives innovation for the library sector.

Chen Jianlong, Director of Peking Universities Library, highlighted libraries' key role in forging a scientific base based on books, and noting past challenges, such as the introduction of foreign language books and of the 'digital library' in 2012.

A formal awards ceremony followed where thanks and gifts were presented to various donors of books and financial contributions to Fudan Library.

Gift being presented to a donor

Symposium on Scientific Data Management and Services in Universities

In the afternoon, a special symposium was held on the challenge of research data management and data sharing in China, and specifically, how libraries could rise to meet these. In the room were invited experts on emerging technologies for open access, such as blockchain, as well as the leading librarians from across China.

A panel of senior librarians discussed issues, with live questions being taken using WeChat, a social media platform. Questions raised included collection usability, size, and the role of data editing; all topics familiar to the UK Data Service.
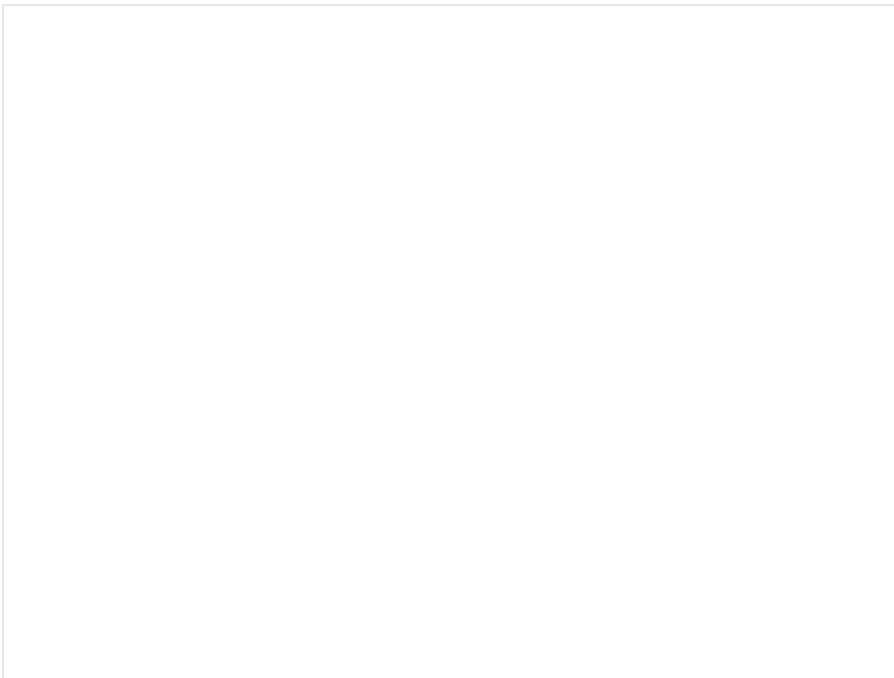

Picture: WeChat questions posed by the audience

My talk was on managing and sharing data in social science, focusing on more historical data.

'Think before Digitising' was welcomed, where cataloguing and box-listing are done first, followed by digitising on demand, preferably with scholars funding the work.

Linking a scholar to a collection before making it digital is a good strategy, as was done for the UK Data Service **Digital Futures (https://ukdataservicedigitalfutures.wordpress.com/)** project.

My intro slide

Following my talk, we heard from Zhang Jing, deputy director of Fudan University Library, and executive deputy director of the Institute of Humanities and Social Sciences, on the position of humanities and social science data at Fudan University.

He recounted that this year China's Ministry on Science and Technology has published positive mandates on FAIR, GDPR and Scientific Data Management Regulations as national policies**.**

Prestige universities, such as Fudan, have embraced this mandate, demonstrating that they can be quickly responsive to higher-level demands.

The new **Scientific Data Research Institute**, hosted by the Fudan Development Institute (FDDI) (http://fddi.en.fudan.edu.cn/) brings together a multidisciplinary mix of big data experts, prominent policy analysts and library scientists. The Contemporary China Social Life Data and Research Center, a Big Data Centre, enables – largely economists – to access real-time population and economic data via their secure-access Hadoop system. Working with government and companies, their 'NextGeneration' Research Platform supports large and federated data storage, sharing and analysis with role-based authorisation. The big data platform makes available around five hundred datasets on energy flow and carbon use, while smaller research datasets deposited in their local Dataverse.

This project has thought carefully about the pipelines required to support different disciplines for data evaluation, processing, and analysis. Data tools include existing commonly used tools by social scientists (SPSS, Stata SAS, WinMax etc.) as well as 3D visualisation tools and an article-publishing platform. The platform is underpinned by new big data platform tools like Spark, Kafka, VPS Docker, similar to the approach being used by the UK Data Service's emerging **Data Services as Platform** (http://dsaap.info/) **(**DSaaP).
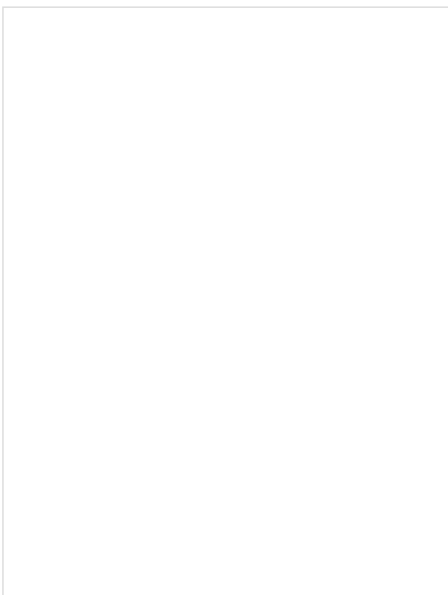

Picture: UK Data Service Data Services as a Platform (DSaaP)

Following the Party Secretary General's request for librarians to quickly skill up on data, many university leaders have mandated that their libraries should take the lead on establishing data management and repository services. It was noted that while the Ministry has great expectations, it is less aware of the scope and scale of the undertaking and the current lack of expertise. University-level data policies were felt to be a good start; in the same way that UK progressed the landscape, when in 2014 the EPSRC set the bar by mandating University-level data sharing responsibilities.

Further discussion pointed to the new roles and skills required for librarians' jobs and the need for information, data and statistical literacy to be fully recognised and included in librarian training. Evolution of the profession to support new policies is welcomed by librarians across China, who argue that they need to be better integrated with IT specialists in various disciplines.

Before platforms are built, training and advocacy in data management and sharing need to come; and adding these into core training is one way to facilitate this. Topics of openness, privacy, incentives to share, and data re-use are relatively new to researchers in China. Our Sage Handbook, *Managing and Sharing Research Data*, translated into Chinese, can support this, and is already being used in in library science training.

Sage Publications, Managing and Sharing Research Data: A guide to good practice, Corti, van den Eynden, Bishop and Woollard

Indexing in China

Fudan also hosted the annual international conference of 'Indexing Societies', taken here to mean bespoke and predominantly indexing of published text.

Many of the participants I met were experienced independent contractors who indexed a massive range of publications, from PhDs and biographies to indigenous materials and cookbooks to scientific high school textbooks. 'Time-tested' manual efforts were used to index using dedicated commercial software, and one could train to be an indexer by taking a 6-month accredited course.

During the conference I learned of China's previous position on the role of indexing and use of scholarly material. It was explained to me that in the past China has not recognised the need to 'index' their scholarly materials, including books. Further, culturally, scholars have not needed to cite scholarly information, as published knowledge is deemed to be free and available to all. This is a markedly different situation to those of us used to dealing with citation, accreditation, intellectual property and publishers' copyright.

I found out that the various indexing softwares on the market, including CINDEX (http://www.indexres.com/), MACREX (http://www.macrex.com/) and SKY Index (http://www.sky-software.com/), are used for the indexer to manually create index entries directly in the software, with the software then organising them into alphabetical or page number order. These can then be imported as an RTF into a publishers' own system. A Chinese participant and I both enquired about the use of XML for more effective and efficient indexing, as well as import/export, but sadly this didn't lead to any further discussion around reusable ontologies or schemas.

I presented on the case of social science data as a new kind of FAIR (Findable Accessible Interoperable Reusable) output/publication that needed some degree of indexing to improve its FAIRness. For example, key words can capture method and study coverage, and the content of questions and code books.

I raised the question of how best to index qualitative data – the research method (the questions asked) or by the content of the data (answers)? I hope to follow this up with an article in the Journal of Indexing on this conundrum, with support from my expert colleagues at the UK Data Service.

Autoindexing: a solution or a sin?

Surprisingly, questions about using automated or semi-automated methods employing Machine Learning and AI for indexing books, especially of more complex or scientific nature, were not well received; the chair suggest that, 'If a machine can replace me, I won't have job'. She argued that the subtleties of language in a text such as a biography, where co-referencing is commonly used (e.g. Louise Corti, the Associate Director, her etc.) could not be easily indexed, displaying her lack of awareness of the sophisticated algorithms used in Natural Language Processing to successfully identify co-references.

Having just worked with my colleague, Jeannine Beeken (https://data-archive.ac.uk/about/staff?sid=jb16483) on a recent CLARIN workshop (https://www.clarin.eu/blog/oral-history-under-scrutiny-m%C3%BCnchen-cross-disciplinary-overtures-between-linguists-historians) where we introduced linguistic approaches and tools for oral history data, I advised them to try out the free Stanford CoreNLP (http://nlp.stanford.edu:8080/corenlp/) to try out named entity recognition and coreferencing.

Curious? Try this for yourself. Copy the italic text and paste into the search box in the demo version (http://nlp.stanford.edu:8080/corenlp/): *Louise is a Service Director at the UK Data Service. She has been working there for 30 years. She often enjoys her role, especially when she gets invited to China*

Further, a tale was recounted – more than once – of their community being promised an autoindexing system by IBM some 20 years ago, which never materialised. I was somewhat disappointed in this fear of the machine, yet it renewed my own aim to get the UK Data Service question and variable autoindexing project started soon!

Postscript

Many thanks to my host for the week, data librarian, Yin Shenquin who, together with her colleagues, have undertaken progressive work in China to set up a small network of data centres in Chinese academic libraries (populated Dataverses (https://www.slideshare.net/datascienceiqss/shenqin-dataverse-in-china-v10)). She was also responsible for the Chinese translation of our Sage Handbook, *Managing and Sharing Research Data* . This is being well-received by library science professionals and master's students.

I really enjoyed my invitation to Shanghai and was honoured to be part of Fudan Library's hundred-year celebrations. The need for academic libraries to rapidly transform to support a connected digital era is certainly recognised by the leading libraries across China. They have been bold and reached out to colleagues in the progressive data and technology world to partner in developing new information and data platforms.

Louise Corti is Service Director, Collections Development and Producer Relations for the UK Data Service.

Louise leads two teams dedicated to enriching the breadth and quality of the UK Data Service data collection: Collections Development and Producer Support. She is an associate director of the UK Data Archive with special expertise in research data management and the archiving and reuse of qualitative data.

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

Post Comment

Categories

Select Category

Posts by month

Select Month

Search the Data Impact blog

Search   Search

- UK Data Service (https://www.ukdataservice.ac.uk)
- Our impact (https://impact.ukdataservice.ac.uk/)
- Impact Case Studies (https://beta.ukdataservice.ac.uk/impact/case-studies)

UK Data Service

- Data Impact Fellows
  (http://blog.ukdataservice.ac.uk/category/data-impact-fellows/)
- Impact and Innovation Lab
  (http://blog.ukdataservice.ac.uk/category/impact-and-
  innovation-lab/)
- Accessibility (http://blog.ukdataservice.ac.uk/accessibility/)

Funded by the Economic and Social Research Council (ESRC), the UK Data Service provides access to international social, economic and population data.

The Data Impact Blog

The Data Impact blog is run by the UK Data Service and is a hub for anyone interested in maximising the impact of social, population and economic data in research and policy.

Latest Post

Using data for effective health service evaluation: why we need data stewards (http://blog.mimas.ac.uk/ukdataservice/?p=5238)