

COLLABORATIVE MECHANISM FOR PUBLIC DIGITAL PRESERVATION SERVICE IN CHINA

Wu Zhenxin

*National Science Library,
Chinese Academy of Sciences
Department of Library
Information and Archive
Sciences, University of
Chinese Academy of Sciences
wuzx@mail.las.ac.cn*

Zhang Xiaolin

*National Science Library,
Chinese Academy of Sciences
Department of Library
Information and Archive
Sciences, University of
Chinese Academy of Sciences
zhangxl@mail.las.ac.cn*

Zheng Jiancheng

*National Science Library,
Chinese Academy of Sciences
zhengjc@mail.las.ac.cn*

Zhang Dongrong

*National Science Library,
Chinese Academy of Sciences
Department of Library
Information and Archive
Sciences, University of
Chinese Academy of Sciences
zhangdr@mail.las.ac.cn*

Fu Honghu

*National Science Library,
Chinese Academy of Sciences
fuhh@mail.las.ac.cn*

Abstract: National Digital Preservation Program is built as a public service of, by, and for the national library community, and a trustworthy partner in the chain of digital scholarly communications. Its inherent collaboration is built on three layers: a national collaborative commitment by the library community and the government enforced by community auditing and certification, a multi-stakeholder contractual participation in the preservation of targeted resources, and a cooperative network of qualified archiving nodes from and answering to the communities, together with reliable partnerships with international publishers. By designing collaboration into the operation of the service, it has achieved a powerful, robust, trusted, and sustainable digital preservation service for the Chinese library community.

Keywords: digital preservation, collaboration, library, NDPP, China

I. INTRODUCTION

We are in a world where “born digital” is already mainstream in scholarly communications. Hence the sustainable access to the digital scholarly resources is now a pressing strategic issue for any national R&D and education systems.

However, the sustainability of those resources cannot be taken for granted, due to the inherent risks of digital information in terms of media, formats, software, and systems, and because of the international and commercial nature of the scholarly publishing market. This is even so for China which is now the No.1 contributor of scientific publications[1] but most of its international publications are by 20 commercial publishers operated outside China[2].

Some of Chinese institutions started digital preservation early on, such as the preservation of SpringerLink publications at the National Science Library (NSL), Chinese Academy of Science (CAS), in 2009[3], and following preservation of resources such as IOPP publications[4]. But it was keenly realized that to make it successful for digital preservation services in line with the needs of a country like China, it had to be a national collaboration to make use of the collective strengths of the whole library community which has been segmented into many a few library consortia and their different channels of funding. So NSL decided

around 2013 that it would turn its own digital preservation service over into a national one, run by the National Science and Technology Library (NSTL), funded by the Ministry of Science and Technology (MoST), whose legal duty is as the national platform to ensure access to scientific information. The key challenge then was to design a mechanism of service that is supported, participated, and trusted by the national library community and preservation partners in the publishing industry. This paper describes the principles and implementations of its inherent collaboration mechanism.

II. TRENDS OF COLLABORATION IN DIGITAL PRESERVATION IN THE WORLD

As a risk prevention effort itself, the key to digital preservation is its trustworthiness and sustainability. However, technical, economical, legal, organizational, and personnel challenges in achieving these usually demand collaboration among and across communities, to build shared values, to develop standards and best practices, to mobilize resources, to foster trusted relationship and operations, just name a few.

In the progress of digital preservation, collaboration at the national level has played an important role. For example, NDIIPP[5] has established a collaborated preservation mechanism to develop tools, systems, digital infrastructural services, for nation-wide preservation of digital resources. The Netherlands Coalition for Digital Preservation (NCDD) has formed shared national infrastructure through cross-domain cooperation[6]. The cooperation in preservation between the German national specialized libraries[7], the Finland CSC-IT Center for Science' effort for extensive cross-field cooperation with archives, libraries and museums[8], the Canadian Government Information Digital Preservation Network (CGI DPN) work on a geographically dispersed infrastructure that collects and preserves all government information[9], and British Library's working together with intentional digital preservation communities to develop and implement best practices[10], are all examples.

Meanwhile, collaboration at international level promoted and enabled many large scale operations. The most influential examples are Portico[11] and CLOCKSS[12] where libraries and publishers work hand in hand to ensure the substantial coverage of digital publications and the sustainable

preservation of them. In a wider sense, collaboration extends into knowledge sharing and capacity building, such as DPC[13], DCC[14], Nestor[15], OPF[16], IIPC[17], etc. Some are on specific topics, such as Presto PRIME[18], MetaArchive[19], and NIBG[20], still some others are for specific technologies or tools, such as E-ARK[21], ArcLight[22], and Bit Curator[23].

III. COLLABORATIVE OPERATING PRINCIPLES OF NDPP

NSTL[24] is a virtual library set up in 2000 to ensure access to and service of international scientific literature. Its 9 founding members are all national STEM information institutions, including NSL. It originally focused on cooperated acquisition of printed journals and proceedings with a division of labour among its members and provided mainly union catalog, online retrieval, and document delivery services for all the R&E communities in the country. It later gradually embarked on acquisition of retrospective digital publications and small publisher resources for the nation. It is funded by the Ministry of Finance and MoST, and supervised by a Board of Governors with members from the research and library communities.

When NSL turned the digital preservation service to NSTL, the aim was not to shield off any of NSL responsibility or financial burden, but to make sure that community "ownership", support, trust, monitoring will be in place to guarantee a trustworthy and sustainable preservation service of, by, and for the library community and the research and education communities they serve. Several key points were intensively discussed, which later operationalized into the implementation of the new National Digital Preservation Program (NDPP) [25].

(1) The need to mobilize the full force of the entire national acquisition market where library systems are segmented with funding from different jurisdictions, to archive localized archiving within the mainland China as required by laws and regulations.

(2) The need to achieve economy at scale for digital preservation and to avoid substantial add-on expenses for individual libraries already financially challenged.

(3) The need to ensure effectiveness in a field with complicated methods, changing technologies, long term operations, and requirements for

robustness and reliability, demanding a capacity well beyond any one or group of regular libraries.

(4) The need to build trust between the preservation service provider and the library community, and between the provider and publishers, and to have the trust embedded in the operations of the services and verifiable in a participatory and transparent way.

The consensus based on the discussion manifested into the Four Principles of NDPP: Government Investment, Collaborative Program, Trusted Repositories, and Public Service.

Government Investment: Digital preservation of scholarly STEM resources is a strategic public need, so the Government has a legal responsibility to invest and to ensure its sustainability. "Government investment" also serves as a strong stimulator to the libraries to participate with a strong sense of public good and without worrying about the expenses. It, in the meantime, acts as a high-level endorser and guarantor to the publishers for its public status and trust-based operation.

Collabrative Program: NDPP builds on collaboration in multiple dimesions, first as a common vision represented by the Joint Statement on Long-term Preservation of Digital Publication Resources[26], entrusted national preservation agreements with publishers[27] that are authorized by the library community, a network of archiving nodes from representative library systems, and a participatory auditing and certification mechanism to put NDPP under the scrutiny of the community.

Trusted Repositories: Digital preservation and collaboration both live on trust, and the trustworthiness goes beyond technology and system. It has to be based on commonly accepted standards and practices, operationalized, measurable and verifiable, transparent, and accountable. The trustworthiness has to be proven by standard audit and certification.

Public Service: NDPP will not charge its participants, library or publisher, for the preservation. A subscribing library for a preserved resource will get free services when, under a trigger event, the resource is turned on to search and retrieval by its subscribing institusions. A publisher for a resource will not be responsible for processing and preservation cost as long as it provides the data

in community accepted practices. This alleviates the burden for participation.

Concrete mechanisms were designed to implement the principles into operational details. The following sections will give details on some of them.

IV. PARTICIPATORY STRUCTURE OF NDPP

Figure 1 illustrates the participatory structure of NDPP, where each type of participant plays a distinguished but yet interrelated role.

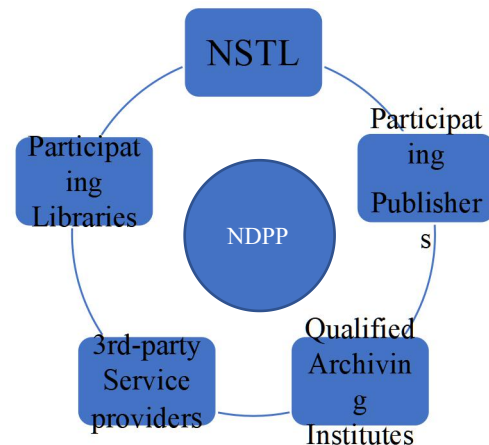


Figure 1 Participatory structure of NDPP

NSTL is the main investor and oversight agency on behalf of the MoST, with annual planning, budgeting, approving administrative or operational procedures, execution monitoring, and assessment of NDPP. It also joins NDPP audit and certification as a "special observer". Members of its Board of Governors serve also as community representatives to oversee NDPP operations.

Participating Libraries are represented by the 200+ major libraries or consoirtia signing the Joint Statement. They review annual reports from NDPP, join in the national demand of availability of the China-based local preservation of the resources they subscrib, authorize NDPP to preserve their subscribed resources, and participate in remote auditing of the resources and on-site certification of the archiving nodes of NDPP.

Participating Publishers are active partners, recognizing that being preservation-friendly is good for business and NDPP and its participating libraries are offering a service to their continuous success in China. They take part in the preservation service not just by signing preservation agreements, but also by providing standards and best practices, working

together with NDPP to solve problems throughout the preservation lifecycle, and ensuring swift changes to NDPP-based services when triggered.

Third party service providers are those responsible for contracted services, such as standards development, tools development, audit and certification organization, system security assessment, format registries, tool registries, keepers registries, and other community or market-oriented services that are necessary for robustness and sustainability of NDPP. It has been an intentional effort for NDPP to entrust some of its key services to proven specialized or infrastructural providers for better efficiency and reliability.

The network of qualified archiving institutes is the operational core of NDPP. They will be introduced in detail in the next section.

V. NETWORK OF QUALIFIED ARCHIVING INSTITUTES OF NDPP[28]

The archiving operation of NDPP is purposely organized in a distributed and collaborated way, even at the expenses of duplicating the tech repository and enduring the overhead cost of managing more than one node.

NDPP's current three nodes include NSL, representing national and quasi-national research academies, Institute of Scientific and Technologic Information of China (ISTIC), representing public STEM information institutes, and Peking University Library (PKUL), representing academic libraries. While each acting as the operator of an archiving repository with a planned division of labour for difference publisher resources, each is also responsible to persuade and unite their peers in respective communities to support NDPP's nationwide efforts. Additional ernational STEM publishers and to ensure the reliability and sustainability of its services.

To become an operating node of NDPP, an institute has to be eligible as a Qualified Archiving Institute (QAI). In addition to being a publicly funded institute, the qualification includes: a substantial range of subscribed scholarly resources, an in-depth familiarity with selection, acquisition, use, and monitor digital publications, a commitment to conformance to the standards and procedures and practices of NDPP, a proven team with necessary administrative and tech capability for a complicate system, a willingness to support public audit and

certification, responsibility for the cost for the physical infrastructure of the archiving system, and a willingness to provide financial assistance when there are needs for it. And, all the QAIs need to abide by the planning, reporting, and assessment guidelines of NSTL and NDPP, many of the reports are published openly on the NDPP website. A formal contract with a QAI would be signed before it is entrusted and invested to run the node.

At the writing of this paper, NDPP, through its three QAIs, has preserved 68 scientific and technological literature databases, including major publishing houses such as Elsevier, Springer Nature, Wiley, Taylor & Francis, and many major societies and university publishers such as IOP, RSC, AIP, OUP, CUP, etc. All together, there are 11,939 international journals, 134,600 academic monographs, more than 834,000 dissertations, 27,092 domestic journals, and 1.8 million+ pre-prints literature, and 20+ retrospective databases of international STEM databases.

The number of QAIs is intentionally kept small to enable tech and financial efficiency, but additional ones could be added under the condition that they bring in unique resources or representing unique communities.

The succession of QAIs is not without consideration, because risk prevention includes the rare case that the QAI itself runs into existential risks. It is in the agreements with publishers and QAIs that the preserved resources would be transferred to another QAI with the consent of both the publisher and the two QAIs concerned.

VI. THE COORDINATED OPERATION OF NDPP

The underlining preservation system was developed following the OAI framework and standards for functions such ingest process, preservation management, public services, and coordinating services which are modeled after the Keepers Registry[29].

The general coordinated operation of NDPP can be described in Fig.2. The three QAIs operate respectively with the same software to run a complete preservation service system for the digital resources it is responsible to preserve. At the same time, they register into the management platform and synchronize their ingest information and metadata regularly. The cooperative management enables the management node to register and

monitor the progress of preservation services at each node and at any time. This structure provides flexibility for future inclusion of additional QAIs.

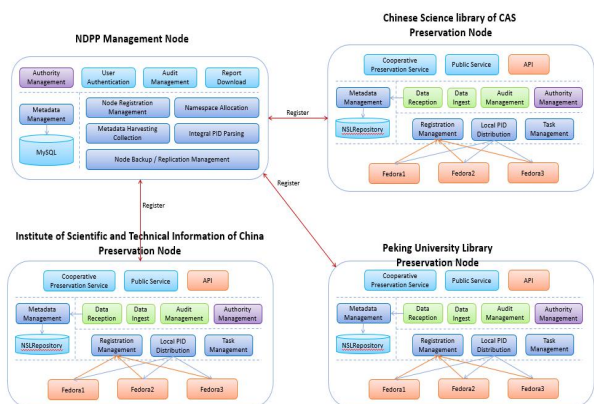


Figure Coordinated Operation Management of NDPP[30]

VII. OPEN AUDITING AND CERTIFICATION OF NDPP OPERATIONS[31]

The trustworthiness of NDPP needs to be measured and verified to ensure long term participation of the library community. This measurement and verification process needs to be inclusive and transparent. From its beginning, NDPP has implemented an open and participatory way of doing this.

Annual remote audits are conducted by selected participating libraries of individual resources at QAIs. At each audit, NDPP invites 10 libraries other than the three QAIs to audit one selected resource, say Springer journals. A randomization tool is provided so the auditor can randomly select certain journals from the resource, certain issues from the journals, certain articles from the issues, to see if they are preserved, and to compare the preserved articles in real time with their online copies directly from the publisher site. Multiple resources are selected each year for each of the QAIs to be audited. The results are compiled independently to be reported to NDPP and NSTL.

Annual on-site audit and certification visits are performed, with teams consisting of representative participating libraries, experts from the IT and information system fields, and responsible NSTL/NDPP personnel, to visit QAIs. Auditing of resources selected is conducted on-site, and certification reviews examine the entire process of digital preservation in accordance with the mainstream international standards, such as the European framework for audit and certification[32],

CoreTrustSeal[33], NESTOR Catalogue[34], ISO 16363:2012[35]. The results will be reported to NDPP and NSTL, and eventually to the library community.

These efforts would help NDPP to prevent the severe problems in preserving what it is said to preserve, whether caused by system failure, technical errors, negligence in management, environmental risks, or publisher mishappens. Doing it in an open way adds **neede** pressure to the QAIs and NDPP itself, but certainly encourages collaboration from the community.

In addition to the audit and certification, a Community Report mechanism[36] is being tested in that NDPP publishes a community report regularly, to communicate with the participating libraries and publishers on the progress of NDPP, problems in preservation, efforts in progress, and challenges lie ahead. It is not as formal as the audit & certification reports but it adds an enlightened way of cooperation.

Currently, NDPP has initiated a project to develop The Guidelines and Regulations System for Trustworthiness Conformance and Its Auditing[37]. It will cover the trustworthy requirements and ways to audit their conformance in risk analysis, strategic planning, annual operational planning, rights management, ingest processing, preservation processing, preservation tech and systems, public services when triggered, system safety and security, back-up operation, succession operation, qualified archiving institute management, financial management, personnel management, audit and certification, and conformance management. The development will be accelerated by partnering with publishers who have a natural interest in trustworthiness and eager to participate in this regard. The preliminary results will be put to consultation with the participating libraries before they are finally made publicly available, so to lay down a stronger base for trusted collaboration.

VIII. FUTURE WORK

As a national preservation service, we understand the importance of preserving digital scholarly resources. The work ahead is to utilize fuller and better the power, resource, experience, and creativity of the community in an inherently inclusive, participatory, democratic, and effective way of collaboration, even failure-proof against ill-

performance or intention of any single player, including NDPP itself.

What makes this more urgent is the coming of age of the new normal of the digital world with increasingly enriching digital objects and resources, digitally-based professional operations (like digital research or digital learning or digital culture), and even digitally-enabled lifecycles and lives. Though NDPP has its origin at digital publications, it cannot ignore the paradigmatic changes of digital objects created, linked, used and processed by research and education communities and their partners. Digital preservation is extending into many unexplored and unexpected fields, and it has to go where researchers, educators, and students go, even if it directs to unfamiliar and untested territories. This requires more, more sophisticated, and more technologically and managerially smart collaboration. One aspect of this will be international cooperation systematically built in the design and operation of NDPP, and embedded into functionalities and processes of its tech platform and its trustworthiness conformance system.

IX. REFERENCES

- [1] Publications Output: U.S. Trends and International Comparisons <https://nces.nsf.gov/pubs/nsb20206/publication-output-by-region-country-or-economy>
- [2] ESAC Market Watch. <https://esac-initiative.org/market-watch/>
- [3] Signing of digital preservation agreement between National Science Library and Springer, 21 Spt.2009 (in Chinese) http://www.las.cas.cn/xwzx/zyxw/200909/t20090908_2468333.html
- [4] National Science Library signed long term preservation agreement with IOPP (in Chinese). http://www.las.cas.cn/xwzx/zyxw/201009/t20100916_2964744.html
- [5] The National Digital Information Infrastructure and Preservation Program. <https://digitalpreservation.gov/>
- [6] Joost van der Nat, Constructing a network of nationwide facilities together. <https://www.dpconline.org/docs/miscellaneous/events/2017-events/1640-dpc-webinar-jvdn-slides-25-jan-2017/file>
- [7] Zarnitz, M., Bähr, T. and Arning, U., 2019. Ten Years of Strategic Collaboration of Libraries in Digital Preservation. LIBER Quarterly, 29(1), pp.1-22
- [8] Johan Kylander, Heikki Helin, Kimmo Koivunen, et al. Together Forever, or How We Created a Common and Collaborative Digital Preservation Service. https://ipres2019.org/static/pdf/iPres2019_paper_117.pdf
- [9] Amanda Wakaruk, Keeping government information on the record: stewardship in a digital era. <https://era.library.ualberta.ca/items/6c1ec79e-a2a4-4212-b9a0-0d11adfb6502/view/8d4d8dad-212b-4058-864b-b4aac0fbb795/CARL-20Leviathan-202015-20Wakaruk.pdf>
- [10] Digital preservation: Saving our digital heritage for future generations. <https://www.bl.uk/digital-preservation>
- [11] Portico. <https://www.portico.org/>
- [12] CLOCKSS. <https://clockss.org/>
- [13] Digital Preservation Coalition. <https://www.dpconline.org>
- [14] Digital Curation Centre. <https://www.dcc.ac.uk/>
- [15] NESTOR. https://www.dnb.de/EN/Professionell/ProjekteKooperationen/nestor/nestor_node.html
- [16] Open Preservation Foundation. <https://openpreservation.org/>
- [17] International Internet Preservation Coalition. <https://netpreserve.org/>
- [18] PrestoPRIME. <https://pro.europeana.eu/project/prestoprime>
- [19] MetaArchive. <https://metaarchive.org/>
- [20] Ana van Meegen Silva. One repository solution doesn't fit all. Towards a distributed and collaborative approach to digital preservation services at the Amsterdam City Archives. https://ipres2019.org/static/pdf/iPres2019_paper_25.pdf
- [21] E-ARK. <https://e-ark-project.com/>
- [22] ArcLight. <https://library.stanford.edu/projects/arclight/>
- [23] Bit Curator. <https://bitcurator.net/>
- [24] National Science and Technology Library. <https://www.nstl.gov.cn/>
- [25] National Digital Preservation Program. <http://ndpp.ac.cn/>
- [26] Joint Statement on Long-term Preservation of Digital Publication Resources http://www.ndpp.ac.cn/jrwm/hzsm/202007/t20200713_568205.html
- [27] National Science Library signed digital preservation agreement with Springer and DRAA. http://www.las.cas.cn/xwzx/zyxw/201105/t20110519_3137216.html
- [28] Authorizing Specification for The Qualified Archiving Institutes of NDPP. 2015, modified 2020. Internal guideline.
- [29] Keepers Registry. <https://keepers.issn.org/>
- [30] Wu Zhenxin, Fu Honghu. Research on the Construction of Distributed Cooperative Preservation Network of Digital Information Resources. Digital Library Forum, 2016(9):43-48.
- [31] Audit Specification of NDPP for Digital Preservation System Trustworthiness. 2015, modified 2020. Internal guideline.
- [32] European Framework for Audit and Certification of Digital Repositories. <http://www.trusteddigitalrepository.eu/>
- [33] CoreTrustSeal. <https://www.coretrustseal.org/>
- [34] Catalogue of Criteria for Trusted Digital Repositories. https://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf
- [35] ISO 16363:2012 - Space data and information transfer systems – Audit and certification of trustworthy digital repositories. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510
- [36] Community Report procedure by NDPP. 2020. <http://ndpp.ac.cn/>
- [37] The Guidelines and Regulations System for Trustworthiness Conformance and Its Auditing for NDPP. 2021.2. Internal Document

