Research on Development Trends of Preservation Metadata

Authors: Liu Jianhua, Zhang Zhixiong, Liu Jianhua

Date: 2016-06-12T00:00:00+00:00

Abstract

In long-term digital preservation, preservation metadata constitutes critical information that supports the long-term viability, renderability, understandability, authenticity, and integrity of digital resources. To comprehensively understand the latest developments in preservation metadata and to provide references for domestic institutions engaged in long-term preservation to formulate implementation strategies and construction plans for preservation metadata, the author comprehensively analyzed the research progress of core theoretical standards, important conferences, and key projects in the long-term preservation field, and summarized the significant development trends of preservation metadata in areas such as rights metadata, significant properties metadata, preservation metadata for special types of digital resources and disciplinary characteristics, semanticization of preservation metadata, and provenance metadata.

Full Text

Preamble

Study on the Trend of Preservation Metadata

Liu Jianhua, Zhang Zhixiong (National Science Library, Chinese Academy of Sciences)

Abstract

In long-term digital preservation, preservation metadata serves as crucial information supporting the long-term viability, renderability, understandability, authenticity, and integrity of digital resources. To comprehensively understand the latest developments in preservation metadata and provide references for domestic institutions engaged in long-term preservation to formulate implementation strategies and construction plans, the authors comprehensively analyzed research progress in core theoretical standards, important conferences, and key

projects in the preservation field, summarizing important development trends in preservation metadata regarding rights metadata, significant properties metadata, preservation metadata for special types of digital resources and disciplinary characteristics, semantic enrichment of preservation metadata, and provenance metadata.

KEY WORDS: Preservation Metadata, PREMIS, Rights Metadata, Significant Properties, Provenance Metadata

1 Overview of Preservation Metadata

Preservation metadata has consistently been a key focus and hot topic in the field of long-term digital resource preservation. As information that supports the long-term preservation process[1], it is essential for maintaining the long-term viability, renderability, and understandability of digital resources. Thanks to preservation metadata, digital objects can become self-documenting, ensuring their long-term preservation and accessibility even when ownership, custody, technology, legal restrictions, or user communities change.

After decades of research, preservation metadata has evolved through three stages: conceptual development, theoretical development, and practical validation. Internationally, numerous mature preservation metadata frameworks, usage methods, and corresponding tools have been developed. Although domestic efforts have primarily focused on application, several standard specification application guidelines have also emerged[2].

As preservation environments continuously change, resource types diversify, and preservation requirements adjust, preservation metadata continues to evolve. This paper aims to explore the development of preservation metadata, analyze its trends based on current theoretical and practical research progress, and provide references for formulating implementation strategies and construction plans for long-term preservation metadata.

2 Development of Preservation Metadata

Research on preservation metadata in the long-term preservation field emerged in the 1990s. As preservation activities expanded, researchers realized that achieving long-term viability, renderability, understandability, authenticity, and consistency for digital resources required recording extensive information about the resources themselves, their hardware and software environments, and change histories. Based on this understanding, they began exploring metadata sets specifically designed for preservation.

This paper is a research outcome of the National Social Science Fund project "Research and Practice on Long-term Digital Preservation Technology" (Approval No.: 09FTQ005).

Corresponding author: Liu Jianhua, Email: liujh@mail.las.ac.cn

In 1998, Michael Day[3] first adopted the concept of "Preservation Metadata" in his report. In the same year, the Research Library Group (RLG) explored a corresponding long-term preservation metadata set for digital image preservation[4]. Subsequently, the Consultative Committee for Space Data Systems (CCSDS) published the Reference Model for an Open Archival Information System (OAIS), providing a reference model and basic conceptual framework for various long-term preservation activities. The OAIS information model proposes that an information object in a digital preservation system consists of the data object itself and its representation information. A data object comprises one or more bit sequences, while representation information converts these bit sequences into more meaningful information. Representation information can be divided into structural representation information and semantic representation information: the former indicates the composition structure of the data object, while the latter expresses its semantic meaning. One representation information can reference already-defined representation information. Typically, to make a preserved information object "independently understandable," it requires a network of representation information composed of a series of referencing relationships. Based on content and function, OAIS further classifies information objects in digital preservation systems into four types: Content Information, Preservation Description Information, Packaging Information, and Descriptive Information[5], explicitly stating that Descriptive Information includes reference information, environment information, provenance information, fixity information, and access rights information (with access rights information added in the 2009 OAIS revision). The UK CURL Exemplars in Digital ARchiveS (CEDARS)[6] project developed a preservation metadata framework for various digital formats by fully absorbing the OAIS reference model and metadata research results from other fields. Additionally, the National Library of Australia[7] and the European Networked European Deposit Library (NetLib)[8] also proposed metadata frameworks for preservation needs during this stage. The publication of OAIS, CEDARS, NetLib, and others marked the transition of preservation metadata from concept to theoretical standard.

In 2000, the Online Computer Library Center (OCLC) and RLG jointly launched an action plan to research the infrastructure of digital preservation metadata[1], continuing to use the term "Preservation Metadata." They developed a broadly applicable, comprehensive metadata framework and published a data dictionary for practical operations. In 2005, PREMIS 1.0[9] was officially released, followed by PREMIS 2.0, 2.1, 2.2[10], and 2.3 over the next decade. Currently, PREMIS 3.0 is under preparation but has not been officially released. The official release of PREMIS upgraded preservation metadata from theoretical standards to practical operations. Since PREMIS, research on preservation metadata has become increasingly application-oriented, achieving considerable results in specific implementation strategies and corresponding metadata extraction tools.

In addition to international research outcomes, domestic contributions have been primarily represented by explorations from the National Library of China,

Tsinghua University, and military academies in preservation metadata frameworks. Under the support of the "National Digital Library Engineering Long-term Preservation Standards" project, Cheng Bian' ai and colleagues from Tsinghua University published the "Application Guide for National Digital Library Long-term Preservation Metadata Standards" [2]. Based on standards and the PREMIS data dictionary, this guide lists essential semantic units, identifier namespaces and controlled vocabularies required for preservation systems, and introduces several tools for automatic metadata extraction along with practical scenarios, providing excellent guidance for implementation.

3 Trends and Implementation Recommendations for Preservation Metadata

Along with the deepening of digital preservation theoretical research and practical activities, and considering factors such as practical usability, the need for clear rights information in collaborative and distributed preservation environments, diversified preservation content requirements, multiple preservation strategy needs, and dynamic preservation demands, numerous institutions and projects engaged in long-term preservation research have pioneered new developments in preservation metadata from both theoretical and practical perspectives. Through comprehensive analysis of research progress in core standards, important conferences, and key projects in the long-term preservation field, the authors believe that preservation metadata has shown significant development in rights metadata, significant properties metadata, preservation metadata for special types of digital resources and disciplinary characteristics, semantic enrichment of preservation metadata, and provenance metadata. Based on this analysis, implementation recommendations are provided. The following sections will elaborate on these aspects individually.

3.1 Increasing Emphasis on Rights Information

As distributed and collaborative preservation gradually become mainstream in long-term preservation, copyright exclusivity may restrict or even completely hinder the preservation of works through copying and dissemination. Simultaneously, post-preservation operations increasingly require more rights information. Consequently, rights metadata has become a major focus within the preservation metadata system.

The 2009 revision of OAIS explicitly distinguished a category of access rights information under its existing Preservation Description Information (PDI). This information primarily determines access constraints related to content information, including legal frameworks, licensing terms, and access controls. It contains access and distribution conditions under compliance with submission agreements, along with descriptions of rights enforcement measures. This marked a clear theoretical foundation for rights metadata as an important component of preservation metadata. In 2013, the Digital Preservation Coalition (DPC)

published *Preservation Metadata (2nd edition)*[11], which explicitly categorized preservation metadata into three types: provenance metadata, rights metadata, and technical/environmental descriptive metadata. These developments conceptually and theoretically emphasize the importance of rights metadata to preservation metadata.

From a practical perspective, PREMIS[9-10] has made rights metadata adjustments and changes the absolute primary focus in its three most recent metadata dictionary versions. These include expanding rights metadata semantic units, extending the previously single "license statement" to three explicit forms of intellectual property rights: copyright, license, and statute, while also providing an extensible semantic unit for intellectual property that cannot be classified into these three forms. Additionally, the new PREMIS versions provide both generic metadata and proprietary metadata for the expanded rights statement containers, such as identifiers, characteristics, categories, and rights features granted to repositories, greatly enriching the definition of rights metadata. Furthermore, PREMIS has further enriched the definitions of relationships between rights metadata and other entities.

The Personal Archives Accessible in Digital Media (Paradigm) project[12], a collaboration between Oxford University and the University of Manchester, treats rights information as an important component of long-term preservation metadata, focusing on intellectual property rights metadata. They believe intellectual property rights metadata is significant for both preservation and access. Based on extensive examination of various rights metadata types, Paradigm established a rights preservation scheme for personal archives. In Rosetta[13], a long-term preservation solution jointly developed by Ex Libris and the National Library of New Zealand, rights management serves as one of the core preservation metadata components, used both for preserving rights metadata and controlling access to preserved objects. The PrestoPRIME project[14], funded by the EU Seventh Framework Programme and focusing on audiovisual materials, designed a detailed rights management metadata ontology model based on the Media Value Chain Ontology (ISO/IEC 21000-19) after thoroughly investigating existing rights expression languages and formats such as MPEG-21, Open Digital Rights Language, and PREMIS, as well as rights management systems like Media Streaming MAF and MPEG Extensible Middleware, to support the long-term preservation of audiovisual resources[15].

Overall, whether independent rights metadata schemes or rights metadata embedded within preservation metadata systems, definitions for rights metadata have matured. Domestic preservation institutions should at minimum record authorized preservation actions that the preservation system can perform on preserved digital objects during implementation. For additional rights description information oriented toward resource services, appropriate semantic units should be selected for recording based on actual needs.

3.2 Increasingly Specific and Clear Metadata for Digital Resource Significant Properties

In preservation metadata, the concept of Significant Properties is one of the most debated topics in the digital preservation metadata community, directly influencing preservation strategy selection in repositories. Researchers generally agree that significant properties are not inherent to digital objects but are key characteristics (such as appearance, look-and-feel, intellectual content, etc.) that must be maintained and are determined by the environment in which preservation activities occur as time passes. However, previously, "significant properties" had been given various definitions and used in many different ways, lacking consistent agreement on how to categorize and use them. In 2009, David Giaretta from the CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project[16] discussed significant properties, authenticity, provenance, representation information, and OAIS at the iPres 2009 conference. Building upon definitions proposed in the new OAIS version, he clarified the relationships between significant properties, authenticity, provenance, and representation information, providing application examples that offered strong theoretical support for using significant properties. The PREMIS 2.0 Data Dictionary, officially published in 2008, created a structured set of semantic units for significant properties. These semantic units facilitate the management of significant properties of preserved objects while ensuring these properties remain undamaged by preservation activities, providing operational metadata support for implementing significant properties. To thoroughly investigate the complete concept of significant properties, identify properties important for various types of preservation objects, and evaluate the importance of each property for future object rendering, the Joint Information Systems Committee (JISC) specifically funded the Investigating the Significant Properties of Electronic Content Over Time (InSPECT) project[17].

According to its research objectives, the InSPECT project collected and organized various definitions and viewpoints on significant properties, analyzed and developed a clear and consistent concept[18], and subsequently formed a complete Significant Properties Data Dictionary[19] that clearly defines semantic units related to significant properties. The project' s research results provide important references on significant properties for JISC-funded long-term preservation projects. Additionally, Simone Sacchi et al.[20] proposed a logical model for significant properties in the long-term preservation of scientific datasets. This model provides the necessary foundation for identifying and assigning significant properties to appropriate entities. Although no long-term preservation project has yet developed a vocabulary for this metadata, some institutions have begun applying it in practice to record preservation information. For example, the UK' s SHERPA Digital Humanities Data Service preservation project constructed significant properties metadata for e-print resources to record special attributes such as semantic content (text or images) and document layout, facilitating future resource maintenance[2]. Chinese insti-

tutions should also gradually emphasize this type of metadata during long-term preservation implementation, referring to OAIS, PREMIS, InSPECT, and other projects to develop semantic units that comprehensively describe object properties from a resource utilization perspective, and record corresponding significant properties metadata during storage.

3.3 Increasing Attention to Preservation Metadata for Special Types of Digital Resources and Disciplinary Characteristics

As long-term preservation continues to develop, preservation work that originally focused mainly on digital text resources has gradually expanded to include more resource types, such as audio/video files, interactive multimedia resources, scientific data (e.g., satellite remote sensing data in earth sciences), web content (webpages, blogs, etc.), interactive content, applications, and processes. Meanwhile, long-term preservation work that was previously concentrated in cultural institutions such as libraries, archives, and museums for traditional cultural resources has gradually expanded into various disciplinary fields. In this context, a single preservation metadata system cannot meet the preservation needs of diverse digital resource types. Therefore, many researchers have designed corresponding preservation metadata models for digital resources with disciplinary characteristics or special types by combining the features of the resources to be preserved with disciplinary needs and leveraging existing metadata systems and domain knowledge structures.

Ruth Duerr et al.[21] designed a long-term preservation metadata model for scientific data in the earth sciences domain based on metadata systems including PREMIS, ISO 1115, and the Content Standard for Digital Geospatial Metadata (CSDGM).

Kia Ng et al.[22-23] proposed the CASPAR ontology metadata framework, which can effectively describe interactive multimedia resources, by referencing and reusing OAIS, the CIDOC Conceptual Reference Model (CIDOC-CRM), and the FRBR model. Günter R. Fuhr et al.[24] utilized existing formal ontologies in the domain to define and describe preservation metadata for digital objects in medical institutions and clinical research data. The Presto-PRIME project[14], funded by the EU Seventh Framework Programme and focusing on audiovisual materials, needed to address the long-term preservation of born-digital video materials. To achieve this goal, the project expanded the scope of preservation metadata based on OAIS, PREMIS, and Rosetta preservation metadata, incorporating provenance metadata, technical metadata, and rights metadata as important components. The EU BlogForever project[25] specifically targeted social network data such as blogs, designing a preservation metadata model that includes blog content, blog feeds, network and linked data, classification content, content semantic information, spam detection information, harvesting information, external plugins, ranking, and other elements, providing excellent references for the long-term preservation, management, and use of blog resources.

As domestic preservation activities gradually expand, current long-term preservation work primarily focused on digital text resources will inevitably extend to other types, while disciplinary preservation for specific fields will also increase. In response, institutions with adequate resources can develop complete, dedicated long-term preservation metadata models based on existing mature metadata systems. However, most institutions can fully utilize extension mechanisms provided in existing metadata systems, such as PREMIS' s extension semantic unit mechanism, to embed externally-defined metadata for special formats into corresponding semantic component containers, such as the NISO Z39.87 MIX schema for digital still images. This approach enables rapid extension of metadata for different types and disciplines.

3.4 Increasingly Rich Semantics in Preservation Metadata

With the development of semantic technologies, preservation metadata has increasingly emphasized the integration and application of semantic technologies. Compared with traditional simple metadata descriptions, the introduction of semantic technologies can facilitate deeper revelation of digital resource content by preservation metadata, enhance the depth of content mining, and further improve the performance of preservation metadata in terms of granularity and interoperability.

The enrichment of preservation metadata semantics is achieved through two aspects: its own development and implementation processes. In its own development, preservation metadata primarily applies the OWL language and uses RDF format for storage and management. For example, PREMIS released an OWL ontology in June 2013 that provides a data dictionary compatible with linked data and supports PREMIS serialization[26].

In practical implementation, semantic enrichment is mainly reflected in the application of semantic technologies across different stages. Specifically, this includes using RDF triples to describe preservation metadata records and applying linked data for digital resource organization. The EU-funded Memories project[27] established RDF triples among preservation metadata to describe classes, subclasses, properties, terms, and relationships, forming an ontology-based preservation structure for long-term audio resource preservation. The French SPAR project[28] utilized the METS standard to establish connections between METS and RDF triples, for example, using <infor:bnf/spar/provenance#hasEvent> to correspond to the "hasEvent" property in the provenance ontology. Beyond RDF triples, the SPAR project also attempted to integrate the entire data repository as linked data. The SCIDIP-ES project (Science Data Infrastructure for Preservation-Earth Science), funded by the EU FP7 Framework Programme[29], investigated and analyzed technical issues related to earth science data preservation metadata, concluding that the application of ontology resources such as "GeoNames" and the General Multilingual Environmental Thesaurus (GEMET) improved the description and preservation of geospatial data by preservation metadata.

Kia Ng et al.[22-23] referenced and reused existing relevant standard models such as CIDOC-CRM and the FRBR model to propose the CASPAR ontology metadata framework, which describes the complex relationships among various concepts in interactive multimedia resources, enabling an ontology-driven preservation approach for interactive multimedia resources. Through this framework, they described relationships and dependencies among preservation metadata more comprehensively and effectively. Based on this research, they argued that combining ontologies with preservation metadata in long-term preservation applications greatly benefits digital repository interoperability and automatic reasoning of relationships among digital objects. Günter R. Fuhr et al.[24] proposed a preservation metadata system for digital objects in medical institutions and clinical research data based on existing domain ontologies. This metadata system is also expressed in ontology form, primarily describing basic characteristics of data objects (such as type, format, size, preservation description information, etc.) and some domain-specific information. With this domain ontology, object indexing and metadata management can be achieved, thereby resolving usage issues for medical domain users caused by format obsolescence and changes in medical terminology. The EU SHAMAN (Sustaining Heritage Access) context model provides an independent infrastructure using ontologies to represent digital object attributes and their relationships[30], advancing the application of semantic technologies in long-term preservation.

From the aforementioned research, it is evident that current implementation of semantics primarily relies on three technologies: RDF triples, linked data, and ontology frameworks, with ontologies being predominant. Therefore, Chinese preservation institutions can fully draw upon and apply existing general or disciplinary ontology frameworks, using ontologies to describe, store, and manage metadata when conducting relevant preservation work.

3.5 Increasingly Rich and Refined Provenance Metadata

Provenance metadata is an important category of preservation metadata responsible for recording information about actions and their results during digital object creation and preservation processes, along with some verification information. This metadata is crucial for ensuring the authenticity of preserved objects[31]. With continuous development of digital preservation practices and the Semantic Web, provenance metadata has become increasingly rich and refined in both content and expression models. Regarding content, some preservation repositories believe digital provenance should include complete historical records of all changes since object creation, including migrations, normalization, and other operations. In PREMIS, objects are immutable; all modification operations generate a new object record that establishes a relationship with its source object. For all objects, provenance-related metadata includes the object' s creator, owner, rights holder, events or process operations affecting the object, and operation times. These metadata elements are represented through corresponding semantic units under PREMIS event entities and agent entities, while

object entities also provide extension elements including object characteristics or significant properties (including domain-specific provenance vocabularies). The National Library of New Zealand (NLNZ) defines provenance metadata to include process descriptions. These elements are similar to events in PREMIS but involve multiple steps over longer time periods in NLNZ, which also provides metadata for recording the reasons for process implementation. Ex Libris' s Rosetta explicitly identifies provenance metadata as one of the core elements in its preservation metadata scheme. This metadata records the origin of digital objects and all changes occurring on them, i.e., records of series of events, actions, and related actors. METS also provides specific elements named digiProvMD in its administrative metadata section to describe any preservation activities related to digital objects. Regarding provenance metadata expression models, the Open Provenance Model (OPM)[32], released in 2007, is most representative. OPM aims to define an interoperable provenance information model across different systems, allowing developers to create and share tools that manipulate this model. OPM defines provenance from a technical perspective, supporting provenance descriptions for multiple entities and allowing multi-level descriptions to coexist. The model defines three core concepts—Artifact, Process, and Agent—and six relationships and roles. In this model, time can also be annotated for reference. Other models, such as Provenir Ontology[33] and Provenance Vocabulary[34], have also proposed provenance metadata representation models from different perspectives.

Based on this development trend, preservation implementation institutions should fully consider recording provenance information during digital resource preservation, especially for scientific data. Under the current circumstances without a unified universal model, implementations can select one existing expression model, such as OPM, Provenir, or Provenance, as a general foundation. Based on this foundation, personalized provenance models can be constructed according to the institution' s preservation strategies or the disciplinary domain of preserved content, defining appropriate granularity for provenance information description to facilitate recording relevant provenance information.

As crucial information supporting the long-term digital preservation process, preservation metadata is essential for maintaining the long-term viability, renderability, and understandability of digital resources. As preservation environments continuously change, resource types diversify, and preservation requirements adjust, preservation metadata continues to evolve. This paper combines current theoretical and practical research progress in preservation metadata to analyze its development trends. This analysis serves two purposes: first, to comprehensively understand the latest developments in preservation metadata, and second, to provide references for domestic institutions in formulating implementation strategies and construction plans for long-term preservation metadata, supporting them in designing suitable preservation metadata collections based on their respective preservation characteristics.

References

[1] Preservation Metadata for Digital Objects: A Review of the State of the Art, A White Paper by the OCLC/RLG Working Group on Preservation Metadata[OL].[2014-10-13]. http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

[2] 程变爱，郑小惠，童庆钧，等. 国家数字图书馆长期保存元数据标准规范应用指南 [K]. [2014-10-13].http://www.nlc.gov.cn/newstgc/gjsztsggc/bzgf/201101/W020120412526555524627.

[3] Michael Day.Issues and Approaches to Preservation Metadata[C].In: Guidelines for digital imaging: papers given at the joint National Preservation Office and Research Libraries Group preservation conference in Warwick, 1998. National Preservation Office, London, pp. 73-84.

[4] RLG Working Group on Preservation Issues of Metadata[OL].[2014-12-13]. http://www.rlg.org/preserv/presmeta.html

[5] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS) [OL]. CCSDS 650.0-B-1 BLUE BOOK.Washington, DC. [2014-10-13]. http://public.ccsds.org/publications/archive/650x0b1s.pdf

[6] Michael Day.CEDARS:Digital Preservation and Metadata[C]. In: 6th DELOS Workshop: Preservation of Digital Information, 1998, Tomar

[7] National Library of Australia.Preservation Metadata for Digital Collections-Exposure Draft [OL].[2014-12-13].http://pandora.nla.gov.au/pan/25498/20020625-0000/www.nla.gov.au/preserve/pmeta.html

[8] Titia van der Werf-Davelaar. Long-term Preservation of Electronic Publications-The NEDLIB projec[J].D-Lib Magazine,1999(5):9

[9] Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group [OL]. [2014-12-13]. http://www.oclc.org/research/projects/pmwg/premis-final.pdf

[10] PREMIS Data Dictionary Version 2.2: Hierarchical Listing of Semantic Units[OL].[2014-10-13].http://www.loc.gov/standards/premis/v2/premis-dd-Hierarchical-Listing-2-2.html

[11] Brian Lavoie, Richard Gartner. Preservation Metadata(2nd Edition)[OL].[2014-12-13]. http://dx.doi.org/10.7207/twr13-03

[12] Rights metadata for personal archives. [OL].[2014-12-13]. http://www.paradigm.ac.uk/workbook/metadata
metadata.html

[13] Nir Sherwinter, Richard Wright. Strategy for use of preservation metadata within a digital library[OL].[2014-12-13].https://www.prestocentre.org/system/files/library/resource/strategy_

[14] Guus Schreiber.Metadata Models, Interoperability Gaps and Extensions to Preservation Metadata Standards[OL].[2014-12-13]. https://prestoprimews.ina.fr/public/deliverables/PP_WP

[15] Annarita Di Carlo.Internal Deliverable ID4.0.5b-Common Rights Ontology[OL]. [2014-12-13] . https://prestoprimews.ina.fr/public/deliverables/PP_WP4_ID4.0.5b_RightsOntology

[16] Giaretta David, Matthews Brian, Bicarregui Juan, Lambert Simon,et al. Significant Properties, Authenticity, Provenance, Representation Information and OAIS Information[OL]. [2014-12-13].2009. http://escholarship.org/uc/item/0wf3j9cw

[17] The significant properties of digital objects. [OL].[2014-12-13]. http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops.aspx

[18] Andrew Wilson.Significant Properties Report[OL]. [2014-12-13]. http://www.significantproperties.org.uk/w

[19] Gareth Knight, Digital Curation Specialist .InSPECT Significant Properties Data Dictionary[OL]. [2014-12-14]. http://www.significantproperties.org.uk/sigprop-dictionary.pdf

[20] Simone Sacchi, Karen Wickett, Allen Renear, et al.Framework for Applying the Concept of Significant Properties to Datasets[C]. In: Proceedings of ASIST 2011 , October 9–13, 2011, New Orleans, LA, USA.

[21] Ruth Duerr, Ron Weaver, Mark A. Parsons. Chapter 7-A New Approach to Preservation Metadata for Scientific Data –A RealWorld Example[M],Springer-Verlag Berlin Heidelberg,2010,pp:113-125

[22] Kia Ng, Eleni Mikroyannidi, Bee Ong. An Ontology based Framework for the Preservation of Interactive Multimedia Performances[C].In: Proceedings of 2009 Fifth IEEE International Conference on e-Science,pp:40-44

[23] Tran Vu Pham & Kia Ng. Preservation of Interactive Multimedia Performances: Ontologies and OAIS[OL].[2014-12-13]. http://www.casparpreserves.eu/Members/metaware/Presenta of-interactive-multimedia-performances-ontologies-and-oais/at_download/file.ppt

[24] Günter R. Fuhr, . Heiko Zimmermann. An Ontology Framework for Long-term Digital Preserv ation and beyond[OL].[2014-12-13]. http://www.ibmt.fraunhofer.de/content/dam/ibmt/en/documents/PDFs/ibmt-product-information-sheets/telematics-intelligent-health-systems/MT_ENSURE_en.pdf

[25] K. Stepanyan, M. Joy, A. Cristea, et al.D2.2 Report:BlogForever Data Model[OL]. [2014-12-13]. https://zenodo.org/record/7488/files/BlogForever_D2_2WeblogDataModel.pdf

[26] PREMIS OWL ontology 2.2 now available[OL].[2014-12-13]. http://www.loc.gov/standards/premis/ontolog announcement.html

[27] Jean-François Cosandier, Per Dahl. The European Project MEMORIES: Management, Description, Retrieval of Audio Archives[OL].[2014-12-13]. http://www.memories-project.eu/documents/Presentation_MEMORIES_Amsterdam.ppt

[28] Linked data and preservation metadata[OL].[2014-12-13]. http://www.dpconline.org/component/docman/ linksthatlast2012peyrard

[29] Report on the survey of technologies, polices,metadata,semantics and ontologies[OL].[2014-10-13]. http://www.scidip-es.eu/assets/Deliverables/SCIDIP-ES-DEL-WP15-D15-1.pdf

[30] Kun Qian, Maik Schott, Christian Kraetzer, et al.A Security Contextualisation Framework for Digital Long-Term Preservation. In: Proceedings of

the International Workshop on Semantic Digital Archives, (part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL)), September 29th, 2011, Berlin, Germany.

[31] Priscilla Caplan.DCC Digital Curation Manual: Instalment on Preservation Metadata [OL]. [2014-12-13]. http://hdl.handle.net/1842/3356

[32] The OPM Provenance Model(OPM)[OL].[2014-12-13].http://openprovenance.org

[33] Provenir Ontology[OL].[2014-10-13].http://wiki.knoesis.org/index.php/Provenir_ontology

[34] Hertig O. Provenance Information in the web of Data[C]. In: Proceedings of the Linked Data on the web (LDOW) workshop at WWW, 2009,Madrid, Spain.

Author Information

刘建华女，中国科学院文献情报中心，馆员，长期保存，信息抽取与文本挖掘，E-mail：liujh@mail.las.ac.cn，北京市海淀区，邮编：100190。

张智雄男，中国科学院文献情报中心，研究馆员，博士生导师，信息系统和智能信息处理。地址同上。

Contact Information

通讯地址：北京市海淀区北四环西路 33 号文献情报中心 3D
联系电话：010-82628382；13426465118.

Note: Figure translations are in progress. See original paper for figures.