EXPLORATION OF PRESERVATION METADATA TOWARDS MEDICAL RESOURCES LONG-TERM ARCHIVING

Chenliu Yang
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
yang.chenliu@ imicams.ac.cn

Jiahui Hu
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
hu.jiahui@ imicams.ac.cn

Qian Wang
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
wang.qian@ imicams.ac.cn

Lei Wang
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
wang.lei@ imicams.ac.cn

Kuanda Yao
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
yao.kuanda@ imicams.ac.cn

An Fang
Institute of Medical Information, Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
fang.an@ imicams.ac.cn

Abstract – Preservation metadata is the significant component of digital long-term preservation, includes the information about data processing, and undertakes the functions maintaining the reliability, integrity, authenticity, renderability, understandability of archived resources. This paper describes MedPRES preservation metadata-related work, construction requirements with the foundation of Open Archival Information System (OAIS) architecture, corresponding to PREMIS data model which contains object, event, agent , rights entities, as well as data dictionary, illustrates metadata practice in MedPRES key processes as well.

Keywords – Long Term Preservation, PREMIS, Preservation Metadata, OAIS, Semantic Units

Conference Topics – Enhancing the Collaboration, Building the Capacity & Capability.

I. INTRODUCTION

As the Medical Library of National Science and Technology Library, Institute of Medical Information, Chinese Academy of Medical Sciences (IMICAMS) has already possessed approximately 2,750 thousand literature, subscribed about 4,600 journals and almost 80 databases every year, archived 113.15 terabyte population health data (1370 datasets, 9,690,577 thousand data records). To satisfy the archival demands of these resources, since 2016, IMICAMS started the construction of medical data long-term preservation platform - MedPRES, which aimed to provide archiving services for at least twenty years. As the significant component of MedPRES management, preservation metadata devotes to make sure reliability, integrity, authenticity, availability, and understandability of archived resources. The purpose of this paper is to present MedPRES preservation data model, relations, and dictionary to undertake the obligation of recording the associated technology, methods, tools, history, management mechanism, and illustrate some practical achievements.

Structurally, this paper describes the designing and practices of preservation metadata to the critical processes in MedPRES. Section II illustrates features of medical resources, these characteristics

17th International Conference on Digital Preservation
iPRES 2021, Beijing,China.

DOI: 10.1145/nnnnnnn.nnnnnnn

will be the basic consideration of metadata development. Section III presents the requirements of preservation metadata construction. Section IV introduces the main contents of preservation metadata design, including the basic principles, methods, data model, data dictionary, and logic diagram of the entire design. Section V illustrates the particular metadata practice under the standard and critical functions, assisted to long-term preservation validation and maintenance, and section VI describes the issues and proposal for future research.

II. Features of Medical Resources

Comparing to the general digital resources, medical digital data are unique and special. From the view of characteristics, medical resources are comprised of complex categories, high privacy, strong correlation, and Continuous Archiving features.

A. Complex Categories

There are various medical digital resources, such as measured data (blood pressure, body temperature), text (literature, EMR, EHR), signal (ECG (electrocardio), electroencephalogram (EEG)), image (B-ultrasonic, computerized tomography (CT)), and audio/video (operation video), involved in structured, semi-structured, and unstructured composite data.

B. High Privacy

Medical resources include a mass of private information, especially clinical diagnosis and treatment details, should be protected well not only about ethic but also with legislation. Because of the high privacy, medical resources requires better safeguard than other information property.

C. Highly Connected

Most times, different types of medical resources are related in practice, besides archived original data, the relationships among different materials also should be recorded and preserved. For example, when researchers reuse archived plain text in future, they may demand related image, radio or audio as well.

D. Highly Iterative

Medical research activities and practice usually proceed in several phases, new research achievements are possible the validation, extension or improvements of previous experiments, that is why it is necessary to preserve each stage and results of research. For instance, to evidence-based medical research, it requires reliability, authenticity, integrity and accessibility.

III. Requirements of Metadata

Based on the specific features of medical digital resources, with the goals of long-term archiving and findability, accessibility, interoperability, reusability of medical digital resources, it is essential to distinguish the demands of preservation metadata.

A. Process Recording

Processing environment, resource discovery, and utilization activities are crucial components of preservation metadata, especially how to optimize, administrate the procedures of digital preservation. Metadata works as efficient evidence to trace and audit archiving information and activities, make sure the non-repudiation of implementation. There are several objects in long-term preservation process, and it is important to make out entities, their relations and influence factors.

B. Resource Validation

It is important to prove resource and operation is real, reliable, traceable, and explicable in archiving process. Commonly, annotation and descriptive information (DI) serve as the evidence to check data quality, only successful verified data object will be accepted and preserved. At this time, metadata should be organized and recorded to archive and access the future resources, guarantee the security, accuracy and authenticity of them.

C. Storage Information

Preservation metadata requires to record the storage details in preservations system and data repository, such as data retention, location, relevant software and hardware, so the framework and elements are able to move from the conceptual to practical, and beyond proof of concept, so they will be convinced and efficient on entire data storage processes. For some sensitive resources, it will be essential to make requests of access control and safeguard.

D. Maintenance Standard

Format change and technological obsolescence are big challenges for usability and interpretability. It is feasible to record the associated details for data

format migration, contain a list of versions of designated format, so that users can easily update to the latest version or roll back to any of the old ones. Beyond that, as the significant factor in self-assessment, risk and threat should be shielded and disposed promptly to make sure all preserved items are correct and integrated.

IV. PRESERVATION METADATA DESIGN

With the requirements mentioned above, preservation metadata works as structured records to the whole processes of digital preservation. So it is necessary to stipulate the basic principles to guide, design and create preservation metadata model, approaching to the particular and practical demands of long-term preservation.

A. Principles

1. Rationality

Preservation metadata should meet the needs of each procedure of archiving functional model, comprised of acceptance, ingestion, storage, retrieve, and access of medical digital object.

2. Authority

Program implementation demands to follow the authorized standards, it is requisite to cohere with the current international standards, national standards, and professional standards.

3. Modularization

It is effective to divide preservation metadata elements into several entities, and build the preservation metadata framework by aggregating the different entities together. Besides, entity attributes probably added or deleted depending on the practical situation.

4. Coherence

When describing resource objects, it is efficient to discover the general information from numerous types of digital resources, demonstrate by core elements, and select directed elements on particular resources.

5. Scalability

Along with the preservation demands increase, scalability and modification are important on the relevant elements of preservation metadata model, such as object entities, entity relationships and entity attributes.

B. Standards

Some institutes have already formulated several accredited standards for guiding archiving activities, among them, Preservation Metadata: Implementation Strategies (PREMIS) [1] defined metadata dictionary to record archiving information about digital resources over the long term, and it has the roots in the Open Archival Information System (OAIS) [2] reference model but has been strongly influenced by the practical experience of the repository.

The PREMIS Data Dictionary is a comprehensive, practical resource for implementing preservation metadata in digital preservation system [1]. The PREMIS defines four main entities (Objects, Events, Agents and Rights) and the relationship between metadata Sematic Units of entities and entities through data dictionary. Object is a discrete information unit in digital form; Event refers to the behavior that involves or affects at least one Object or Agent; Agent refers to the individual, organization or software program related to the Events or Rights in the life cycle of the Object; Rights Statement refers to one or more of the Rights Statement of the Object and the Agent. The PREMIS Data Model is shown as Figure 1.

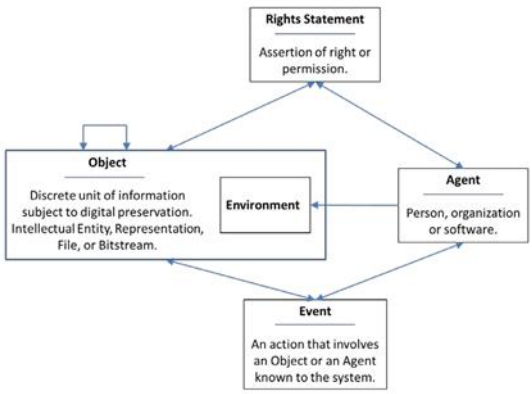

Figure 1: The PREMIS Data Model[1]

The PREMIS 3.0 data dictionary includes a total of 197 semantic units (including containers and independent semantic units), including 90 Object semantic units, 21 Event semantic units, 18 Agent semantic units, and 68 Rights Statement semantic units. PREMIS has provided the metadata that most preservation systems need to define to perform preservation functions under normal circumstances, which has a very good guiding role in the construction of the metadata model of the long-term preservation system. In addition, PREMIS does not consider the description metadata used for

discovery and acquisition, nor does it define the specific format metadata, and it is recommended that when formulating a scheme for storing metadata during implementation, it should be treated according to specific issues.

C. Corresponding

To handle the large amounts of data items with changing technological circumstances, the object should follow the vital metadata which is mandatory for preservation management. What is more, the digital preservation records need to follow the universal standard to make sure future interaction and interpretability.

In the OAIS reference model, representation information, packaging information, and preservation descriptive information (PDI) have remained as the conceptual foundation for the PREMIS data dictionary. However, no matter what entities are involved in the data model, it should include efficient information about the data processing in long-term preservation metadata, and be capable to support archiving decisions, document existed activities, record achievements and effects, ensure the integrity, authenticity, security, and reusability of medical resources simultaneously.

Within archiving process, PREMIS aims to provide sufficient information for taking appropriate measures on the preserved digital objects to maintain and ensure they can be found, used, and interpreted in case of the future technical revolutions.

Based on the four types of entities in PREMIS, identify the corresponding content of the four types of entities in MedPRES.

1) Agent corresponds to the data resource storage institution of MedPRES, and determines the responsible party for data resource storage by assigning a unique identifier to the storage institution, recording the name of the institution, the type of the institution, and other related information. At present, MedPRES has not cooperated with other institutions to provide preservation services, therefore, the archiving system is only MedPRES.

2) Rights corresponds to the rights information stipulated by MedPRES. MedPRES stipulates the preservation rights and preservation contents by signing an agreement which including rights information stipulated with institution. Therefore, a Rights Statement corresponds to an agreement.

3) Object corresponds to the batch and file of MedPRES. Batch corresponds to the list content in the agreement, which is also the general term for the data resources submitted or transmitted in a single time. An agreement list can correspond to one batch or multiple batches; File means that the data resource is stored in the MedPRES's repository in the form of files. A batch may contain one file or multiple files.

4) Event corresponds to the MedPRES event information, which is used to describe the changes in the agent, rights, and object. The event does not consider the containment relationship between batches and files, and only records the changes and related information of different entity objects.

D. Elements Selection

The four entities and semantic units of objects, events, agents, and rights defined by the PREMIS data dictionary basically cover the complete information that needs to be recorded during the long-term preservation of digital resources.

Currently, except for objects entity, MedPRES events, agents, rights entities are relatively simple. Referenced on accredited standards and commonly use situations, select appropriate methods to proceed with metadata design.

1) Reuse PREMIS 3.0 data dictionary are preferential to preservation metadata design. Practically, simplify the semantic units of PREMIS, choose and sustain the better interrelated and pragmatic attributes of medical digital resources.

2) Simplify PREMIS 197 semantic units, select MedPRES practical elements, considering relevance, operability, usability.

3) Some container semantic units in PREMIS data dictionary are modified and directly assigned, so that such semantic units can be valued according to the actual situation.

4) Environmental factors have not been applied in MedPRES for the time being, but MedPRES has started to build a software resource library, and environmental factors will be gradually applied based on actual needs

Data dictionary is comprised of semantic units directed to each entity above. It maps to PREMIS entity properties in repository to some extents, references the necessary semantic units, and provides evidence for the source trace of completed

operation. These records are able to describe the key points on each stage of preservation processes.

1) MedPRES reuses 51 semantic units on describing preservation characters of objects (batch and file), shown as table 1.

Table 1 Object Semantic Units

PREMIS Semantic Units			
objectIdentifie	objectIdentifierType		
	objectIdentifierValue		
objectCategory			
originalName			
preservationLevel	preservationLevelType		
	preservationLevelValue		
significantProperties	significantPropertiesType		
	significantPropertiesValue		
objectCharacteristics	fixity	messageDigestAlgorithm	
		messageDigest	
	size		
	format	formatDesignation	formatName
			formatVersion
		formatRegistry	formatRegistryName
			formatRegistryKey
	creatingApplication	creatingApplicationName	
		creatingApplicationVersion	
	inhibitors	inhibitorType	
		inhibitorTarget	
		inhibitorKey	
storage	contentLocation	contentLocationType	
		contentLocationValue	
	storageMedium		
relationship	relationshipType		
	relationshipSubType		
	relatedObjectIdentifier	relatedObjectIdentifierType	
		relatedObjectIdentifierValue	
		relatedObjectSequence	
	relatedEventIdentifier	relatedEventIdentifierType	
		relatedEventIdentifierValue	
		relatedEventSequence	
linkingEventIdentifier	linkingEventIdentifierType		
	linkingEventIdentifierValue		
linkingRightsStatementIdentifier	linkingRightsStatementIdentifierType		
	linkingRightsStatementIdentifierValue		

*refers to the semantic units used only by files

Batch uses semantic units that describe the resources in agreement, like object identifier, object category, original name, preservation level, significant property, and linking object/event identifier.

File retains the elements of fixity, size, format, creating application, inhibitors, storage, and belonged batch object identifier. For instance, the digest algorithm is the recommended method to check file integrity; format details are used on the latest or future update of indispensable software or application; storage information records the file location in repository. Besides, file object identifier is used on retrieval and management, connecting with related files that belong to the same batch, and complete storage details to some extent.

2) MedPRES reuses 15 semantic units on describing preservation characters of events, shown as table 2.

Table 2 Event Semantic Units

PREMIS Semantic Units	
eventIdentifier	eventIdentifierType
	eventIdentifierValue
eventType	
eventDateTime	
eventDetailInformation	eventDetail
	eventDetailExtension
eventOutcomeInformation	
linkingAgentIdentifier	linkingAgentIdentifierType
	linkingAgentIdentifierValue
linkingObjectIdentifier	linkingObjectIdentifierType
	linkingObjectIdentifierValue

Event comes with the foundation of event identifier, type, outcome, and details, to illustrate partial activities in digital archiving management. In detail, optimize the disposal process records by supplying the particulars of purpose, supervisor, start time, end time, and affected files, guarantee implementation is necessary and efficient, fill out materials of logs, for further audit, and make preservation activities trustworthy.

3) MedPRES reuses 5 semantic units on describing preservation characters of agent, shown as table 3.

Table 3 Agent Semantic Units

PREMIS Semantic Units	
agentIdentifier	agentIdentifierType
	agentIdentifierValue
agentName	
agentType	

Agent plays an important role in the whole process of archiving. With the PREMIS foundation units, agent identifier, name, and type, specify the authorized staff who maintains data object regularly, add contact information for subsequent consultant and mentoring.

4) MedPRES reuses 22 semantic units on describing preservation characters of rights, shown as table 4.

Table 4 Rights Statement Semantic Units

PREMIS Semantic Units			
rightsStatement	rightsStatementIdentifier	rightsStatementIdentifierType	
		rightsStatementIdentifierValue	
	copyrightsInformation	copyrightstatus	
		copyrightsJurisdiction	
	licenseInformation	licenseDocumentationIdentifier	licenseDocumentationIdentifierType
			licenseDocumentationIdentifierValue
		licenseApplicableDates	startDate
			endDate
	rightsGranted	act	
	linkingObjectIdentifier	linkingObjectIdentifierType	
		linkingObjectIdentifierValue	
	linkingAgentIdentifier	linkingAgentIdentifierType	
		linkingAgentIdentifierValue	

Rights statement is connected with agreement between MedPRES and resources depositor, and all details of rights now come from agreement information. MedPRES will get authorization through the protocols.

Besides, the identifier is the most significant element to relationship creation, recording relational and interactional archiving information, preserve relevant files and their relationships for future necessary use and knowledge discovery.

In general, semantic units of *object* and *file* focus on data acceptance and storage procedures, such as data resources confirmation, format registry, and general management; units of the *event* emphasized the change of data object or files, trace executed operation simultaneously, like risk disposition, format migration, or data recovery; units of the *agent* are used to record the responsible archival institute and staff, make sure relevant rights and accountability of archived resources.

V. PRACTICE IN KEY PROCESSES

Then, it is important to employ preservation in key processes of MedPRES, complete and support maintenance, monitoring, planning, conscious actions and strategy implementation [3].

A. Resource Confirmation

Digital archiving includes essential activities to verify and confirm submitted digital objects based on metadata, aiming to make sure digital resources are authentic, refundable, reusable, and interpretable through affiliated annotation or descriptive form filled by the object owner.

MedPRES prepares two precepts of metadata extraction regarding different digital object sources and situations, one is an annotation document for co-operators, and the other is a stated form filled by visitors. Then, annotation or uploaded form will be extracted and generated to verify the descriptive items are integrated and accurate. Some object metadata items will be recorded as part of the preservation metadata.

At the same time, authenticity and integrity will be checked by metadata records before resources are accepted to the preservation agency, and the contents of each item will be retained to describe the archived object. First, make sure the correspondence of file quantity, then, verify the checksum coded with digest algorithm with original value guaranteed by the depositor.

When the batch is accepted to MedPRES successfully, a persistent identifier (PI), batch *objectIdentifier* will be added into metafile, representing object entity within the entire processes, and file(s) of the object will get *file objectIdentifier* after validation. For example, PI is mapping to the corresponding resources in the repository, and keeps sustainable correlations among connected objects, files or events, for future feasible discovery.

With character and archiving requirements, evaluate object *preservation level,* take different measures, and provide appropriate storage conditions for resources on each level. On one hand, it will be useful to safeguard sensitive or confidential resources, on the other hand, it is essential to keep the balance between data sustainable storage and lower cost.

Furthermore, the sponsor or administrator associated with to object cannot be ignored, including organizational viability, staffing, procedural accountability, and preservation policy

framework. Regarding the agent entity, the information of relevant persons, organizations, or systems should be documented and preserved and serve as an "agent" entity with relevant information.

B. Format Registry

Format describes the particular structure of digital object files and presents the relevant applications supporting file interpretability in long-term preservation. Considering the possible sense of technological obsolescence as time goes by, archived digital data files may not be able to read or explain, that is why the relevant information of format should be involved in metadata for subsequent availability.

PRONOM, one of the well-known databases of file formats and technical supports, including hardware, software, and operating systems, holding more than 550 file formats, 250 software products, and 100 vendors [4].

DROID, the format recognition tool of PRONOM, is in charge of verifying the format of digital resources. Then, the identified format will be recorded as authenticated items quoted from PRONOM, shown as the examples in table 2, and the unidentified format will be classified as unauthenticated contents sourced from user supplements in the local database.

Table 2 Examples of Format

Semantic units			MedPRES Example
format	formatDesignation	formatName	Acrobat PDF
		formatVersion	1.0
	formatRegistry	formatRegistryName	PRONOM
		formatRegistryId	fmt/15
		formatRegistryPurpose	validation profile

Format migration and recommendation are significant for digital archiving. The preserved data files should be the same as the originals, so the migration path will be identified and recorded as "Event" attributes in preservation information to provide the relevant tracing services to the object. Besides, when the current format becomes obsolete with the registry update, the system administrator and object owners will receive the alert message at the same time.

With the assistance of preservation information, the critical functions, format registry, and migration are beneficial to make sure the fixity, sustainability, interpretability of preserved contents.

C. Object Management

Metadata is used by the administrative processes to supervise and monitor the relevant activities in object management, playing a vital role in making data findable, accessible, interoperable, and reusable (FAIR). MedPRES works on managing the challenges of entire digital preservation processes and guarantees the digital repository is capable of preserving and archiving resources for a long time. Through the judgment metrics from metadata, the repository completes and supports consistent maintenance, monitoring, planning, as well as conscious actions and strategy implementation.

Firstly, MedPRES provides a global unique identifier to each archived object or file, for findability. Rich metadata indicates the preserved data and the researchable index of whole objects. After validation, object and related files would be given a persistent identifier (PI) refers to batch *objectIdentifier* and *file objectIdentifier* in preservation metadata automatically and persistently. This item will guide the future implementations of storage, management, access, etc.

Then, MedPRES employs a mandatory access control mechanism to files, for accessibility. Unlike commonly digital objects, medical resources have high privacy requirements because they may include personal information or clinic research data, which demands ethical and legal protection. As medical resources archiving system, MedPRES supports conditional data access, archived resources are corresponding to *privilege,* and carry out data or medium encryption technology, so unauthorized organizations or individuals will be restricted to access or disclosure archived resources.

Furthermore, MedPRES devotes to communicating with different languages medical resources based on well-known terms so that they are understandable, for interoperability. No matter the contents or language of digital objects, metadata will store as the proper format, such as the XML and RDF, and the units of metadata are usually formal. Meantime, the relationship among related files will be recorded in preservation information so that users can follow one object or identifier to gather more knowledge around.

Moreover, preservation metadata guarantees the reusability of digital objects, linking resources by RDF in MedPRES. Meanwhile, it enables tracking of the object provenance in the repository, and the *event* entity semantic units will record each implementation associated with targeted objects or files, such as version update and content change. The items of *event* entity are persuasive evidence to supervise all of the implementations to preserved objects and make sure files are available to use time after time, to realize data continuous archiving.

D. Risk Disposition

For digital long-term preservation, the crucial issue is how to avoid the complicated potential threats, so it is necessary to alert and control uncertainties and reduce the risks of archived resources.

Currently, risk analysis enables administrator assessing current and future threats of object formats in the repository, including format change, application obsolescence, and relevant property issues which challenge the sustainability and readability of digital objects. Besides, there are some generic risks to digital preservation [5], like infrastructure failure, data loss, misoperation, backup mistakes, and security issues.

There is no method to guarantee data object is out of risks totally, but previous positive preparation can avoid most problems. Once digital resources archives to the repository, as part of preservation metadata, the contextual or environmental information will indicate to the data object that has already been preserved as well.

Table 3 Storage Semantic Units

Semantic Unit			MedPRES Example
storage	Location (path)		C:\Users\Desktop\DataArchive2020\pack1
	medium	mediumId	H015
		mediumType	magnetic disk
		expiredDate	12-31-2025

Some threats and vulnerabilities cannot be detected or noticed immediately, such as hard disk damage. Shown as table 3, *storage* records file physical location and medium information, real-time monitoring or periodical check values of these items, so that MedPRES can find fault points and solve problems rapidly when storage equipment broke down.

Moreover, the supervision process followed by digital curation is designed to identify, analyze, and evaluate the potential threats and vulnerabilities to the repository based on the contextual and environmental information recording. No matter how to solve the problems, the *event* entity preserves activities and results that involve archived objects or files.

VI. PHASED ACHIEVEMENTS

Shown as figure 3, currently, MedPRES has already built basic functions of data collection, reception, ingestion, preservation, and other related services, archived 6 types of medical resources, including books, journals, webs, science data, tools and information system.

Figure 3 MedPRES Console Interface

Shown as figure 4, preservation records display on archiving medical object detail page, users will be able to view object archiving information and trace the previous implementation.

Figure 4 Medical Object Preservation Detail Page

VII. DISCUSSION

In addition, preservation metadata is persuasive evidence to audit the integrity, consistency, reusability, interpretability. In audit and certification activities, how to select elements and metrics to evaluate pragmatic effect for trustworthy digital

archives is still challenging to preservation metadata further development and improvement. The construction of MedPRES is still ongoing to satisfy and gather the amount of available evidence and implement the audit process [6].

Now, MedPRES just preserved medical literature resources. IMICAMS expects more cooperation with medical resources submitter, analyses the difference and specific characters of medical resources, and attempt to appropriately facilitate and improve current preservation metadata scheme to satisfy the long-term preservation requirements in future.

VIII. CONCLUSION

In this paper, we introduce the design and practice of MedPRES preservation metadata. Based on the medical digital long-term archiving requirements, create the data model which contains object, file, event, agent entities, relevant relationships, and semantic units mapping to each entity. In particular, the contents of semantic units include two parts, one comes from the parallelism elements of PREMIS, and the other relies on the extension units. Meanwhile, we discuss several key practical scenes and activities, such as resource validation, format registry, object management, risk control, audit, and certification, display the current achievements.

From the perspective of sustainability, MedPRES preservation metadata provides operations and solutions which aim to guarantee authenticity, integrity, consistency, interpretability, stability, and sustainability of data objects, making them findable, accessible, interoperable, and reusable.

ACKNOWLEDGMENT

MedPRES is supported by Chinese Academy of Medical Sciences (CAMS) projects, i.e., Research on the Long-term Preservation Strategy of Medical Digital Resources project and Long-Term Preservation of Medical Big Data Construction Equipment Purchase project.

REFERENCES

[1] PREMIS Data Dictionary for Preservation Metadata, 2015. [Online].Available:http://www.dcc.ac.uk/resources/metadata-standards/premis

[2] REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS), The Consultative Committee for Space Data System(CCSDS).[Online].Available:https://www.iso.org/standard/57284.html

[3] AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES, The Consultative Committee for Space Data System(CCSDS).[Online].Available:https://public.ccsds.org/Pubs/652x0m1.pdf

[4] F. Soper. The PRONOM File Format Registry, Experts Workgroup on the Preservation of Digital Memory, 2004. [Online].Available:https://www.erpanet.org/events/workgroup/documents/soper-pronom.pdf

[5] H. M. Gladney, Long-Term Preservation of Digital Records: Trustworthy Digital Objects, THE AMERICAN ARCHIVIST, Vol. 72 (Fall/Winter 2009): 401-435.

[6] Digital Preservation Coalition, Audit and certification, Digital PreservationHandbook,2015.[Online].Available:https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification