

Long-term Preservation Solution for Complex Digital Objects Preserved as Archival Information Packages in the Domain of Pharmaceutical Records

Hrvoje Stancic
Kresimir Pavlina

Department of Information Sciences
Faculty of Humanities and Social Sciences
Zagreb, Croatia
e-mail: hrvoje.stancic@zg.t-com.hr
e-mail: kpavlina@ffzg.hr

Arian Rajh

Agency for Medical Products and Medical Devices
Zagreb, Croatia
e-mail: arian.rajh@halmeh.hr

Abstract—The authors base their research on several standards: Reference Model for an Open Archival Information System (OAIS) ISO 14721 and Information and documentation – Records management processes – Metadata for records standard ISO 2381. They build a model for the long-term preservation of pharmaceutical records in the eCTD file format (electronic Common Technical Document – standard format for pharmaceutical documentation) stored in a digital archive. The model shows formation of archival information packages (AIPs) as structured, complex objects – based on eCTD’s XML elements, packaged together with the appropriate metadata in a single object and protected for the preservation reasons by an integrated checksum creating MDx algorithm. An application based upon the developed model, upon initial mapping of eCTD’s structure categories and metadata to the AIP’s structure, is shown to be able to automate this process as much as possible. This procedure should enable users to use the application even for different types of complex digital objects, not only for eCTD. In that sense the application could become a generic ‘preservation application’ for the AIP creation and long-term preservation. The authors conclude that the developed model, exemplified by the eCTD documentation format and appurtenant application, could improve the AIP creation of pharmaceutical records intended for the long-term preservation in digital archives.

Keywords—OAIS; eCTD; pharmaceutical records; long-term digital preservation; digital archive

I. INTRODUCTION

Electronic pharmaceutical records are complex objects. In the EU (European Union) they usually come in the eCTD (Electronic Common Technical Document) format [4][7]. Each eCTD document consists of a predefined folder/files structure [8]. Such documents are complemented with additional national-level documents and documents produced by the National Competence Authority (NCA). All this documentation together makes a record of one medicine and upon this record it is allowed to be put on the market. This paper will regard these electronic pharmaceutical records as very important and will try to contribute to the formation of actions toward their appropriate long-term preservation.

Long-term preservation of electronic records is, even when simple objects are considered, rather complicated and quite unpredictable process. This is caused by the constant change and advancement of technology – hardware, software, storage media and encoding standards. Bollacker states that “because any single piece of digital media tends to have a relatively short lifetime, we will have to make copies far more often than has been historically required of analogue media” [14]. In turn, the need for constant forward migration influences preserved records. If they are moved from one media and technology to the more advanced ones the authenticity, reliability, integrity, usability [10] and trustworthiness of the records could be endangered. These aspects of preserved pharmaceutical records, consistent with Stamatiadis, as in [1], are very important not only for the business purposes but, more importantly, for the health of patients. Gladney [6] explores the aspect of records’ authenticity stating that “a performance is called authentic if it has integrity and also conforms to a firmly bound, adequately informative, and honest provenance assertion.” Provenance, when once established at the moment of a record’s creation, becomes part of the record thus leaving the integrity problem. This article will focus on solving the problem of records’ integrity within the long-term

preservation environment in the way that the preserved records maintain their properties while the underlying technologies advance since “authenticity ... is closely related to demonstrating the integrity of documents, that is, ensuring that they are complete and unaltered from the time of creation” [3].

The structure of the research presentation in this paper is as follows. Firstly, the relevant standards are described at the level of complexity relevant to the discussion and analysis. Some standards, like OAIS, are considered as well known and therefore their deeper analysis is left out. Next, the concepts and entities of ISO 23081-1 (Information and documentation – Records management processes – Metadata for records) standard is described and mapped to the OAIS’s information packages’ structure. Further, the structure of eCTD is described in detail. The main part of the discussion is around the issue of creation of archival information packages (AIP), exemplified by eCTD based pharmaceutical records. Abstraction of that example leads to the suggestions of AIP creation in general if complex objects are to be preserved. To support this research the application ArhiMed, created by the authors, illustrate the functionalities of an application intended for the long-term preservation of authentic medical records in eCTD format.

II. OAIS RM

Preservation of complex digital objects in a digital archive is best if an archive formation is based upon internationally accepted standards. One of the important ones, although a highly abstract one, is the Reference Model for an Open Archival Information System, known as OAIS RM or ISO 14721 [9]. Since it is a very well known standard only a part of its structure will be described in this place – the one important to create a context for this research.

The OAIS RM consists of three models: Information model, Information package model and Functional model. *Information model* describes Information Object and Information Package. Information Object consists of Data Object (physical or digital) and Representation Information (in the form of Structure Information, Semantic Information or Representation Networks). Figure 1 shows Information Object structure. Information Package is structured into four main components: Content Information (consists of Content Data Object and Representation Information), Preservation Description Information or PDI (consists of Reference, Context, Provenance and Fixity Information), Packaging Information and Descriptive Information. Figure 2 shows Information Package structure. Further, *Information package model* describes three types of packages – Submission (SIP), Archival (AIP) and Dissemination information package (DIP). Figure 3 shows situations in which each of the package types appear. *Functional model* describes functions within a digital archive that need to be developed and their interconnections that need to be established in order for a digital archive to be OAIS compliant. Package concept and its structure from OAIS will be important and often referred to in this paper.

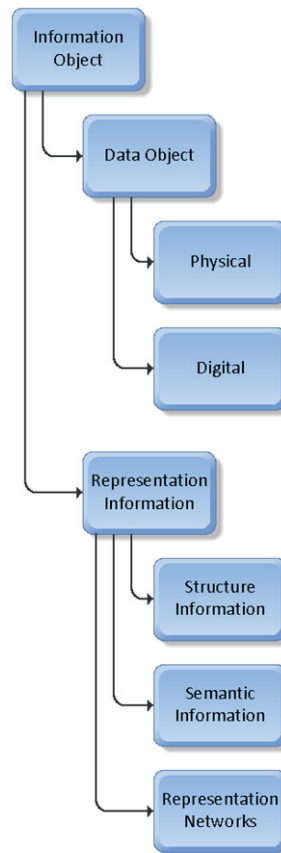


Figure 1. OAIS Information Object structure.

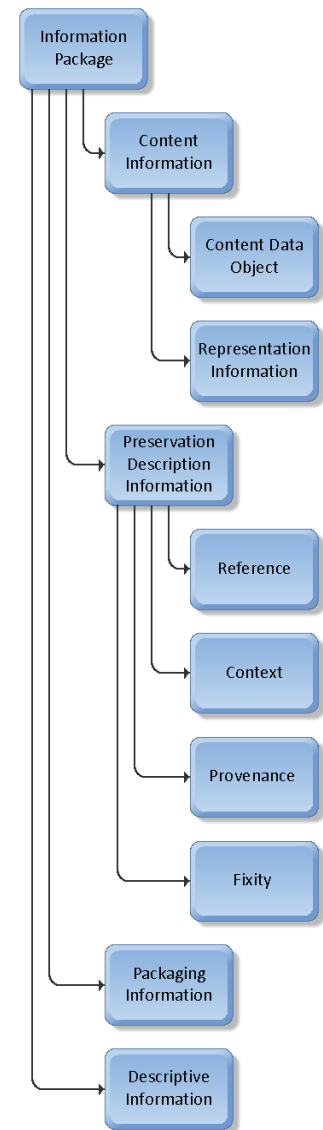


Figure 2. OAIS Information Package structure.

III. ISO 23081

ISO 23081-1:2006 Information and documentation – Records management processes – Metadata for records standard [12] (with part 2 ISO/TS 23081-2:2007 [13]) is addition to the ISO 15489:2001 standard. ISO 15489: 2001 Information and documentation – Records management – Part 1 [10] and Part 2 [11] standardise design and implementation of records management policies, programs and systems. Usage of that standard implies usage of some other standards for specific areas of programs and systems, such as these for work process analysis, quality management, information security, risk management, metadata etc. ISO 23081 standard supplements ISO 15489 by prescribing usage of unambiguous metadata and therefore, by making possible

to use records in different environments, to support interoperability and to disseminate content of recordkeeping systems between various systems and user communities. ISO 23081 standard deals with records' metadata management as a part of records management. Metadata exist to provide records' description but also to decide on records' authenticity (and integrity and reliability in sum), usability and interoperability for users or for other archives. Metadata needs to be captured at the beginning of records lifecycle, because electronic records need to be managed proactively and in advance. Metadata also need to be created after records' submission to a digital archive, i.e., registration, sometimes during their usage and always during automated and semi-automated execution of records management processes and procedures. It is recommended to store produced metadata in metadata database or repository, in the later case especially in the preservation-adequate formats like XML. Stored metadata should be inexorably linked with objects they are referring to, possibly packaged together for the long-term preservation reasons.

The following synthesis of the ISO 23081 standard is intended for better understanding of the importance of metadata and, sometimes, their complexity to either preserve in the present form or migrate to some other metadata standard, inevitably having similar, but never the same structure. ISO 23081 prescribes several *classes and types of metadata* – generally there are metadata on records, metadata on business rules and mandates, metadata on business processes, knowing that records and records management systems are considered to be in metonymic relationship with creator and its business processes, metadata on agents or business process roles, and finally metadata on records management processes and procedures. *Relationships* of these *classes of metadata entities* are as following: agents conduct business, create and use records, mandates govern business, records management subsystem is business-based, and business actions and transactions are interpretable and procurable to agents and other users as records. ISO 23081 provides conceptual model with classes and types of entities. *Class of records* includes aggregated types of records from item, sequence of transaction, file, series, archive and archives – nomenclature is non-enforceable and can be altered in practice to the extent that is necessary. Sequence of transaction is not a sequence of business transaction itself, as the smallest component of business process, but the sequence of physically or virtually linked items evidencing some enclosed business transaction. *Class of agents* includes person or instrument, work group, agency and institution. Persons and working group represent business process role. Business process roles can also be applications and IT subsystems. Applications and systems can be considered as instruments from ISO 23081, but they are specifically defined as roles in business process management discipline, which is an affiliated realm to the ISO 15489 normative area. Institution represent creator in strictly archival sense. *Class of business* includes transaction, activity, function and ambient function. Ambient function is “a societal right or responsibility that exists outside the boundaries of

organization”¹, i.e., creator’s functional place in the wider society. *Class of mandates* consist of business rules as procedural instructions for execution, policies as generic instructions and legislation as legal environment of creator and direct context of his business and archive. ISO 23081 distinguishes *six broad groups of metadata*. Firstly, there are *identity metadata* with elements like entity type (e.g., file), aggregation, and registration identifier (ID of entity in a specific domain). Secondly, there are *descriptive metadata* with elements like title, classification signature, abstract, location information, jurisdictional domain, and external identifiers (if any). Third group is *metadata about usage*, i.e., elements describing technical environment, elements about rights, access/restrictions, intended users, language, integrity (e.g., checksums) and documentary form (structure). Fourth group of metadata, *event plan metadata*, includes elements like event plan (date/time, type of event, description of event, relation) and trigger. *Event history* is the fifth group of ISO 23081 metadata – it consists of identifier, date/time, type, description and relation of event and actions on both the entity and metadata. Sixth group of metadata is *relational metadata* which describes relationships between two or more entities. That relationship becomes a new entity, and its elements are identifiers of related entities, relationship type and relationship date. Metadata should be inherited from classes to types, by linking classes and types, copying metadata from class, or encapsulating aggregates. It should be possible to extract and reuse metadata, add additional elements, and maintain entities as sequences as necessary for the preservation reasons. Metadata support preservation of records but they should be preserved as well.

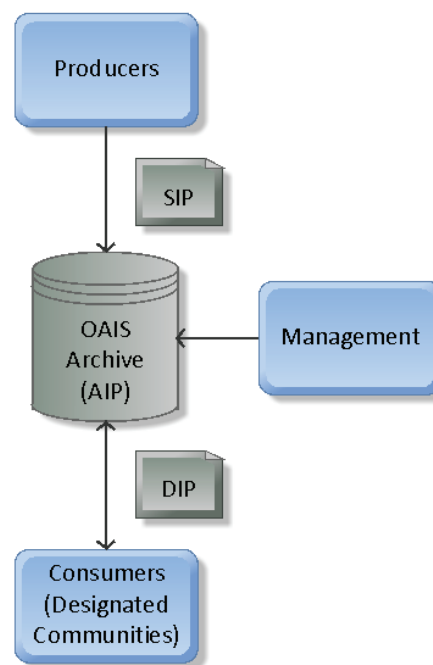


Figure 3. OAIS Information Package model.

¹ ISO/TS 23081-2:2007, Table 2 – Entity class: Business, p. 13.

It is possible and even implied to flatten the conceptual model given in ISO 23081 and to set out metadata entities in different views, e.g., record-centric view with records that interconnect business, mandate and agents, or business-centric view with business entity as nucleus etc. It should also be possible to nest ISO 23081 broad groups of metadata into wider preservation-oriented model, such as OAIS RM, especially into its preservation description information (PDI) and descriptive information – they can be enriched with constructive business-centric and creator-centric ISO 23081 metadata.

IV. ELECTRONIC COMMON TECHNICAL DOCUMENT (ECTD)

In the process of digital preservation sometimes complex digital objects should be considered. One of those objects is the medicinal documentation on a medicine a pharmaceutical company puts on the market. In the process of obtaining a license the medicinal documentation should be submitted for the registration procedure either to the *European medicines agency* (EMA, formally known as EMEA – European medicines evaluation agency) for the centralised EU level procedure or to a national agency (in the case of Croatia it is the Croatian Agency for Medicinal Products and Medical Devices) for the national level registration procedure.

eCTD DTD version 3.2

- m1-administrative-information-and-prescribing-information
 - [EU Module 1](#) [new]
- m2-common-technical-document-summaries
 - m2-3-quality-overall-summary
 - [Quality Expert Statement](#) [new]
 - m2-5-clinical-overview
 - [Clinical Overview](#) [new]
- m3-quality
 - m3-2-body-of-data
 - m3-2-p-drug-product [manufacturer: Big Factory] [product name: WonderPill] [dosage form: Tablets]
 - m3-2-p-1-description-and-composition-of-the-drug-product
 - [Description and Composition of WonderPill](#) [replace]
 - m3-2-p-2-pharmaceutical-development
 - [Pharmaceutical Development](#) [new]
 - m3-2-p-3-manufacture
 - m3-2-p-3-2-batch-formula
 - [Batch Formula](#) [replace]
 - m3-2-p-4-control-of-excipients [excipient: Compendial]
 - [Control of Excipients - Compendial Excipients](#) [replace]
 - m3-2-p-5-control-of-drug-product
 - m3-2-p-5-1-specifications
 - [Specifications](#) [replace]

Figure 4. eCTD content read from the XML backbone.

The documentation submission is obligatory in the processes of registration, re-registration or upgrading of the medicine's documentation. Pharmaceutical companies should submit documentation in the eCTD format [5]. eCTD is basically a set of directories and files grouped in an electronic dossier connected by the XML backbone. Figure 4 shows eCTD example of a fictitious medicine obtained from such XML backbone. Documentation includes administrative documents and summaries, documents on pharmaceutical-chemical quality of a medicine, documents on non-clinical and clinical testing. Parts of the documentation is usually submitted in various file formats, ranging from .pdf, .doc(x), .rtf to .jpeg, .svg, .png, .tif or .gif. The documentation submitted at one point in time, either to the national or international regulatory body, is called a *sequence*. The first sequence's directory is named 0000 (with the strictly defined subdirectory structure), the next one (e.g., for the dosage change) is named 0001 (with its own subdirectory structure repeated), etc. Each sequence contains *index.xml* file (XML backbone) and *index-md5.txt* file². Each sequence consists of *modules* (e.g., m1, m2, m3 etc.). Not all sequence has to have all modules. The most important is module 1 since its *eu* subdirectory contains *eu-regional.xml* file with the information on the medicine. The sequence with the highest number containing module 1 contains the most recent and therefore the most complete information. Figure 5 shows directory structure of a fictitious medicine and figure 6 shows a fictitious example of a medicine's documentation. There are already available commercial solutions for working with eCTD-based documentation and its validity checking, like *EURS is Yours – European Review System* for example.

V. PRESERVATION OF ECTD AS A COMPLEX ELECTRONIC OBJECT

So far the structure of eCTD was described. Nevertheless, eCTD documentation represents only a part of the record on a medicine to be preserved. It can be viewed only as a submission information package (SIP1). In the case of the national medicine registration process the pharmaceutical company has to submit additional documentation specific to the national legislative. Those documents are also organised into a predefined folder/files structure. The folder structure lies beneath the directory named the same as the sequence it refers to with the addition of *workingdocuments* in order to distinguish the two. Figure 7 shows an example of the *workingdocuments* folder additionally describing the 0000 sequence of a fictitious medicine.

This additional documentation is not part of the eCTD and it should be added to it when put in a digital archive (SIP2). Furthermore, there are documents that are created by the National Competence Authority (NCA), a regulatory body, in the process of registration, re-registration or upgrading/changing of the medicine's documentation. For

² Message Digest 5 checksum information intended for the e-dossier (i.e. SIP) integrity check. As it will later be explained, this MD5 information will not be sufficient for the preservation purposes and AIP creation since additional documentation should be added to create a complete record.

each sequence NCA creates documents. Therefore, the complete record of a medicine (see figure 8) consists of:

1. eCTD documentation,
2. *workingdocuments*, and
3. NCA documentation.

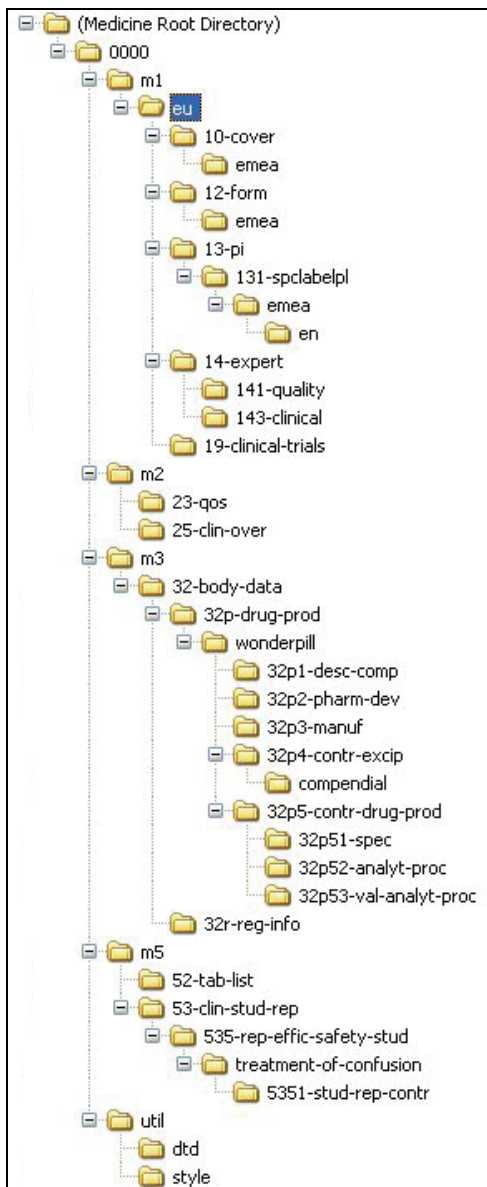


Figure 5. Directory structure of a fictitious medicine.

This is maybe a straightforward process when creating a record, but for the preservation process a digital archivist will encounter:

1. a complex record consisting of multiple folder/files structure,
2. several different file formats, each one in the need of a different preservation method,
3. one part of the record (eCTD) could be validated by MD5 check, while the other two parts can not,

4. some parts of the record are connected by the XML backbone while the others are not, and
5. there are no record-level binder.

EU Module 1

DTD version 1.4

Envelope for EMEA

Submission: Type: Variation Type IA
Mode: Single

Tracking Number(s): EMEA/HC/000123/IA/061

Applicant: PharmaCompany

Agency: EMEA - European Medicines Agency (EU-EMEA)

Procedure: Centralised Procedure

Invented Name: WonderPill

INN: Pioglitazone Hydrochloride

Sequence: 0022

Related Sequence:

Submission Description: Type IA Variation for Change of Address for Big Factory

Module 1 EU

1.0 Cover Letter
For EMEA:

- [Cover Letter for IA/061 \(new\)](#)

1.2 Application Form
For EMEA:

- [Application Form for IA/061 \(new\)](#)
- [Declaration of Change of Address \(new\)](#)

1.3 Product Information

1.3.1 SPC, Labelling and Package Leaflet

1.3.2 Mock-up

1.3.3 Specimen

1.3.4 Consultation with Target Patient Groups

1.3.5 Product Information already approved in the Member

Figure 6. Fictitious example of a medicine's documentation.

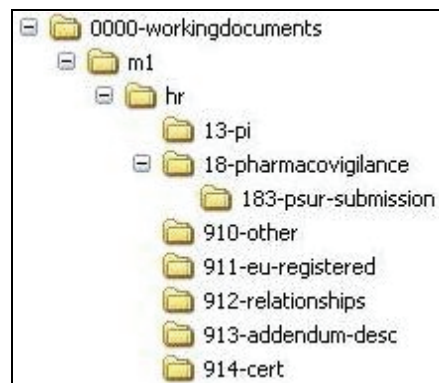


Figure 7. Example of the *workingdocuments* folder.

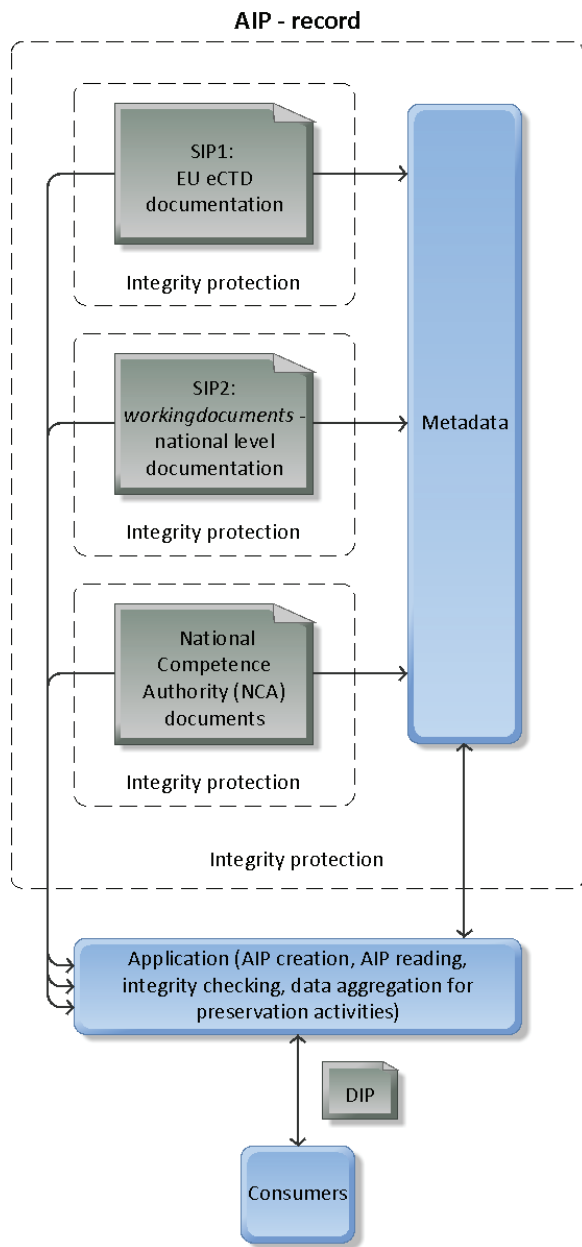


Figure 8. AIP – the complete record of a medicine.

The described situation could be improved by the creation of an application, or building functionality into an already developed digital archive, that can be used for the archival information package (AIP) creation or semi-automatic transformation of SIP into AIP. The functional specification of the application for preservation of pharmaceutical records presented in the following lines could even be made more general in order to be used in a wider spectrum of similar record-preservation situations. The idea underlying the application, which will be shown later in the application functionality description part of this paper, is to integrate all medicine’s documentation parts at the metadata level thus creating one record. It should extract all needed metadata

from the eCTD, enable the preservation person to add additional metadata, automatically add digital archive-specific preservation metadata, take care of the different file formats preserved within the record by triggering certain preservation actions, protect record’s integrity in order to preserve its authenticity, and it should finally create an all-connecting XML file. It would be a very good idea to code and structure metadata according to the internationally accepted standards such as OAIS, ISO 23081 and METS. Having this in mind, an application intended for the creation of a complex electronic object record of a medicine, intended for the long-term preservation in a digital archive, should at least have the following functionalities.

1. The application should be able to extract metadata from the *eu-regional.xml* file. From there it should extract (see figure 6 for this example’s data):

Applicant: PharmaCompany
 Agency: EMEA – European Medicines Agency (EU-EMEA)

Invented Name: WonderPill
 INN: Pioglitazone Hydrochloride

2. From all *m1* modules it should extract submission types of all sequences, e.g.:

Submission: Type: Variation Type IA
 and the DTD version, e.g.:

DTD_Version: 1.4

The submission information represents business procedure (registration, re-registration, upgrading/changing).

The metadata mentioned so far are extracted from the eCTD files. The following archival metadata should either be detected automatically from characteristics of the files to be preserved or manually entered by the preservation specialist.

3. **RecordCreationDate** – date of AIP creation.

4. Person who created AIP – **Agent**: Name, Role (Creator, Editor, Archivist, Preservation, Disseminator, Custodian, IPOwner, Other).

5. **Title** (AIP title = name of the medicine = Invented Name).

6. **AIP_Version** – for distinguishing possible additional sequences and supporting new AIP creation based on prior AIP and newly submitted sequences.

7. **AIP_History** – according to the ISO 23081 this is event history metadata – it records creation dates of all prior AIPs.

8. **RootFolderTitle** – sequences submitted at a later point in time are organised under the same root folder as the prior ones – all being part of the same medicine’s documentation.

9. **External_ID** – optional, free-text field for entering, e.g., NCA’s classification numbers.

10. **FileVersion** – extracts from the folder structure names of all files (**FileName**) and their relative locations, adds information on their **FileTypes** (e.g., pdf, doc, tif etc.) and version (e.g., 1.4). This is important for the later preservation planning process – the information on the precise location of files that need to be migrated could speed up the process.

11. **FileVersionAggregation** – aggregates information on types and versions of all documents and adds their total number in the AIP (e.g., PDF 1.4, 87; doc 2003, 14). This

TABLE I. MAPPING OF METADATA ACROSS DIFFERENT STANDARDS.

Description	Metadata	OAIS	ISO 23081	METS - sections
Who delivers the eCTD package?	Applicant	PDI – Provenance	Agency (identity metadata, entity class: agents, layer: 3)	dmsSec – rightsMD
To whom was initially delivered the eCTD package?	Agency	PDI – Provenance	Institution (identity metadata, entity class: agents, layer: 4)	dmsSec – digiprovMD
What is submitted?	Invented Name	PDI – Reference	title (description metadata; entity class: records, layer: 3)	dmsSec – sourceMD
What is submitted?	INN	PDI – Reference	title (description metadata; entity class: records, layer: 3)	dmsSec – sourceMD
What kind of activity?	Submission	PDI – Context	Activity / process (identity metadata, entity class: business, layer: 2)	dmsSec – digiprovMD
Contents standard	DTD_Version	PDI – Provenance	technical environment (use metadata, entity class: records, layer: 3)	dmsSec – techMD

information supports future preservation action time-frame planning.

12. **Checksum** – it should be created for the whole AIP after its creation.

13. **AIP_Check** – a trigger for AIP’s hash-check and hash-check of all its files (starts automatically after a predefined period of time or manually at any given time).

14. **AIP_VValidityDate** – either the date after which it is possible to (semi-automatically) delete the AIP or information on permanent preservation.

As it was mentioned earlier, the recorded metadata could be mapped to different standards thus enabling interoperability between different digital archives keeping similar records, but also making future preservation activities easier. For example, the metadata extracted from the eCTD files, described through the suggested functionalities 1 and 2, could be mapped across different standards – table 1 (see figure 6 for this example’s data).

VI. ARCHIMED – APPLICATION FOR ARCHIVING AND LONG-TERM PRESERVATION OF ECTD RECORDS

After the initial research in the field of digital preservation and applicable ISO standards, recognition of the potential problems with the long-term preservation, in this case, of medical records, analysis of the eCTD structure, and conception of the model of functionalities for a digital preservation application, we have developed an application called ArchiMed.

Here we will explain and illustrate the basic operating performance of the application. First, ArchiMed is capable of reading all the needed metadata from the eCTD document while enabling the addition of other AIP metadata (Figure 9).

This application form also enables manual input for the hash check interval. It can be set automatically at certain number of days, but can also be changed manually if needed. This example shows the interval set for 30 days. The same semi-automatic approach is applied with AIP validity interval – set for 2 years in this example. AIP validity should be decided upon and set wisely, because after that period AIP data should be examined more closely, i.e., their file formats should be tested for potential migration purposes. For this purposes ArchiMed offers package validation report

which lists name of the medicine, AIP creation date, root folder title, last hash check date and validity period of the AIP (Figure 10).



Figure 9. Creation of the new AIP

At any given moment one can manually initiate data integrity check and test whether any AIP has been either intentionally or unintentionally changed. ArchiMed report shows status of the data integrity check, the total number of packages and the number of those which failed the check (Figure 11).

ArchiMed also offers the possibility of checking which packages will expire in certain number of days. Figure 12, for the purposes of this demonstration, shows those packages which will expire in the next 1,000 days. In practice this feature gives opportunity to a records holding institution to prepare for the potential migration. For example, it can check

which AIPs will expire within, e.g., next 30 days and will need a potential digital preservation action. Along this line of thinking there is another feature in ArchiMed which makes potential migration procedures much easier – file type version report. This report lists all file types appearing in the AIPs (e.g., PDFs), their versions (e.g., 4) and gives the total number of such files. Figure 13 shows one such report. This is very important if it becomes necessary to migrate all PDFs v. 4.1 to the newer format. This report will show if there are any such files and if there are how many of them. It is not the same if you have to migrate 1,000 or 1 million files to the new format. Furthermore, the application can extract only those files needing a preservation action, thus preparing them for the external migration procedure, and later on import new, migrated versions, calculate new hash, and create new, migrated AIP. All this information about the migration procedure is added to the new AIP in order to document the changes. Only in this way, consistent with [2], the authenticity, reliability, integrity and usability of the preserved records could be confirmed and the system for their long-term preservation certified.

VII. CONCLUSION AND FUTURE WORK

The research in this paper showed the complexity of preservation if complex digital objects are to be preserved as authentic records. By using an example from the domain of pharmaceutical records – eCTD documentation – it was shown that the documentation which is submitted to a digital archive in the form of SIP, although structured in a standardised way and complemented by the integrity checking possibilities, still needs to undergo a lot of additional procedures in order to be transformed into an eligible and trusted AIP. In this process the national level documentation is added to the eCTD SIP as well as NCA documents. Only complete documentation can become AIP and stored in a digital archive as an archival record of a medicine's registration procedure. Since the significance of the preservation of the complete registration records in the electronic form for the long-term period needs not to be emphasised, it is of great importance to support such a process in a proactive way. This means to try to preserve as many important metadata about the records themselves, about their context and immediate environment responsible for their correct rendering, functionality and understanding as well as the metadata important for making future preservation actions easier and more efficient, at the point of creating a record. This was a key motivation for this research and analysis which resulted in suggestion of the procedures and functionalities for an application for creating AIPs in the domain of pharmaceutical records, and the creation of the application itself. Currently, ArchiMed is in the process of commercialization. It is planned to become a self-standing, expansion module of a pharmaceutical records management application but it will be possible to obtain it even without the accompanying application.

Future work will be focused on the modification of the model to fit any other type of AIP creating procedure by adding some specific types of metadata or by changing the source XML file to read them from. Coding the metadata

according to several standards, or creating mapping schemes between the standards, makes things a little bit more complicated, but this could, later in the preservation period, prove to be useful. It could also make different designated communities understand archived records better and make search procedures more efficient.

To conclude, the functional application called ArchiMed, built around this specification can produce AIPs securely, quickly and with the zero-error tolerance. By giving aggregated information on files being preserved it can also make migration procedures more easily to plan, test and carry out. Future research and modifications will bring more sophisticated versions having wider functionalities of offering users to customise it to best suite different complex objects' record-creating procedures. By implementing such an approach in the process of long-term preservation, any records holding institution, but also any institution in the pharmaceutical industry, can be sure that their important eCTD or other types of records remain unchanged, and can plan and conduct future migrations without breaching authenticity, reliability, integrity or usability of the preserved records.

REFERENCES

- [1] D. Stamatiadis, "Digital Archiving in the Pharmaceutical Industry," *The Information Management Journal*, pp. 54-59, July/August 2005
- [2] Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), Digital Curation Centre (DCC) and Digital Preservation Europe (DPE) <http://www.repositoryaudit.eu/> 13.10.2010
- [3] E. G. Park, "Understanding 'Authenticity' in Records and Information Management: Analyzing Practitioner Constructs," *The American Archivist*, vol. 64, pp. 270-291, Fall/Winter 2001
- [4] EU Module 1 specification v.1.4, August 2009, <http://esubmission.emea.europa.eu/eumodule1/docs/EU%20M1%201.4/EU%20M1%20Specifications%20v1.4%20FINAL.pdf> 23.12.2010
- [5] Guidance for Industry on Providing Regulatory Information in Electronic Format: eCTD electronic Submissions, Version: 1.0, May 2009, <http://esubmission.emea.europa.eu/doc/eCTD%20Guidance%20Document%201.0%20FINAL%20FOR%20PUBLICATION.pdf> 23.12.2010
- [6] H. M. Gladney, LongTerm Preservation of Digital Records, Part I: A Theoretical Basis, ERPAePRINTS, October 2008, <http://eprints.erpanet.org/152/> 27.12.2010
- [7] ICH M2 EWG electronic common technical document specification, July 2008, http://estri.ich.org/ectd/eCTD_Specification_v3_2_2.pdf 23.12.2010
- [8] ICH M4 harmonised tripartite guideline, Organisation of the common technical document for the registration of pharmaceuticals for human use, January 2004, <http://private.ich.org/LOB/media/MEDIA554.pdf> 23.12.2010
- [9] ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model
- [10] ISO 15489-1 Information and documentation – Records management – Part 1: General (ISO 15489-1:2001)
- [11] ISO 15489-1 Information and documentation – Records management – Part 2: Guidelines (ISO/TR 15489-2:2001)
- [12] ISO 23081-1 Information and documentation – Records management processes – Metadata for records – Part 1: Principles (ISO 23081-1:2006)
- [13] ISO 23081-1 Information and documentation – Records management processes – Metadata for records – Part 2: Conceptual and implementation issues (ISO/TS 23081-2:2007)
- [14] K. D. Bollacker, "Avoiding a Digital Dark Age," *American Scientist*, vol. 98, pp. 106-110, March-April 2010

	Title	Creation Date	Root folder title	Last hash check	Valid until
▶	Demo	25.9.2010.	0022	25.9.2010.	25.9.2012.
	Test	25.9.2010.	0026	25.9.2010.	25.9.2012.
*					

Figure 10. Package validation report

	Title	Creation Date	Last hash check	Check
▶	Demo	25.9.2010.	25.9.2010.	OK
	Test	25.9.2010.	25.9.2010.	Failed
*				

Number of packages:2
Failed data integrity check:1 (50%)

Export Close

Figure 11. Data integrity check report

Show packages that will expire in 1000 day(s) Show

	Title	Creation Date	Root folder title	Last hash check	Valid until
▶	Demo	25.9.2010.	0022	25.9.2010.	25.9.2012.
	Test	25.9.2010.	0026	25.9.2010.	25.9.2012.
*					

Figure 12. Expired packages report

	File Extension	FileType Version	Files
▶	.dtd		4
	.mod		4
	.pdf	4	22
	.bt		2
	.xml		4
	.xsl		4
*			

Figure 13. File type version report