Long-Term Digital Preservation  
for Engineering Designs

|  |  |  |
| --- | --- | --- |
| HMG Services logo.jpg | [**H.M. Gladney**](mailto:gladney@almaden.ibm.com)  Saratoga, CA 95070 |  |

© 2011, H.M. Gladney: These are unpublished notes produced at the private expense of the author. Copies may be freely shared provided that no paragraph is copied for reuse in any document for which HMG is not announced as a co-author.

Any commercial use would be an invasion of copyright unless it has written permission from H.M. Gladney for some specific occasion and purpose agreed to by him.

[Introduction 1](#_Toc308076775)

[Notes 1](#_Toc308076776)

[Changes Needed to TDO Design? 1](#_Toc308076777)

[Extensions Needed to {Regli 2011] 1](#_Toc308076778)

[What Existing Regli Code Can be Exploited? 2](#_Toc308076779)

[Background with Other Peoples‘ Work 2](#_Toc308076780)

[HP & iPedigree 2](#_Toc308076781)

[Requirements 2](#_Toc308076782)

[End-Users’ Specifications 2](#_Toc308076783)

[Software Reusability and Information Portability 3](#_Toc308076784)

[Fundamentals from 20th-Century Epistemology 4](#_Toc308076785)

[Design 4](#_Toc308076786)

[High Level 4](#_Toc308076787)

[Parameter Management 5](#_Toc308076788)

[Design Parameters (for SW Engineers) 5](#_Toc308076789)

[Data Type Parameters Common for All Users 5](#_Toc308076790)

[PDO Structure and Update Management 5](#_Toc308076791)

[Further Considerations 5](#_Toc308076792)

[Conversion Storage-to-Linear and Reverse 5](#_Toc308076793)

[Notes 25th October 5](#_Toc308076794)

[Bibliography 6](#_Toc308076795)

[Links & Current Manuscripts 11](#_Toc308076796)

[*Critiques of other work* 11](#_Toc308076797)

[To Do? 11](#_Toc308076798)

[Candidate Figures 12](#_Toc308076799)

[End Document 15](#_Toc308076800)

[Figure 1: Human and machine roles in sharing documents (*PDI Fig. 2*) 2](#_Toc306509744)

[Figure 2: Human and machine roles in sharing documents (simplified version) 2](#_Toc306509745)

[Figure 3: Model of digital repositories and related human roles (*PDI Fig. 26*) 3](#_Toc306509746)

[Figure 4: Schema for a Trustworthy Digital Object (part of *PDI Fig. 32*) 4](#_Toc306509747)

[Figure 5: Schema for a Preserved Information Package 5](#_Toc306509748)

[Figure 6: Content of a TDO Protection Block (*PDI Fig. 33*) 5](#_Toc306509749)

# 

# Introduction

### Notes

Use [Gladney 2007b] as a starting point (it has not been published).

At the moment, the Bibliography contains article abstracts. These will be suppressed before submission for publication.

The current diagrams are merely a standard set from which we might select.

“PDI” is an acronym for the book Gladney, H.M. [*Preserving Digital Information*](http://www.springer.com/3-540-37886-3), Springer Verlag, 2007.

In order to permit design changes and still have unambiguous referencing, use “TDO” to refer to the object design prior to mid-2001, use “PDO” (“Preservation Digital Object”) to refer to the design about to be implemented, and “PIP” (“Preserved Information Package”) to refer to an abstract design suitable from a proposed ISO standard. 🡺 PIP will be an abstraction of PDO.

Separable challenges:

* Context- and format-independent aspects common to all preserved information objects
* File formats chosen by authors, partly for technical reasons and partly to conform to customary practices of their profession
* Metadata conventions and standards, partly to conform to customary practices of a topic or profession
* Infrastructure to create durable evidence of authenticity

Directly pertinent are [Stancic 2009] and also Regli [2011].   Each responds to needs within a very important market segment whose document formats are specialized, very complex, and bound by standards and industry conventions.  In a good architecture, the solution to (a) and (d) would be sufficient for both these (and for every other information class), whereas the choices within (b) and (c) would be discipline specific.

### Changes Needed to TDO Design?

Ref: [Gladney 2007b], [Gladney 2009], and writings these cite.

1. Does the TDO design need to be changed?
2. Are there widely interesting (generic) extensions to TDO design that we should adopt?
3. Are there TDO extensions that extend to **all** engineering applications?
4. What TDO extensions would cover the entire class of applications considered in [Regli 2009]? What is this class, i.e., can we characterize it abstractly, rather than merely by example?

### Extensions Needed to {Regli 2011]

[Regli 2011] does not sufficiently emphasize the boundary between managing digital objects and managing digital archives.

* Partly this is an issue of philosophical principle.
* Partly it is needed to avoid unsupportable costs.
  + Archives specialized to engineering documents are unaffordable except for very short periods
  + Content needs to be portable to any repository (and any computing environment whatsoever.
  + No single team can provide more than a tiny fraction of the representation tools needed 🡺 the architecture needs to be extensible vis-à-vis file types.

Need to separate issues of ability (tool supported) to preserve from the legal rights to create derivative works.

In a nutshell, there are at least three classes of information for which paper archives were not designed:

Representations of algorithms (including computer programs and engineering procedures);

Representation of mathematical information; and

1. Representation of performances.

A critical challenge is that the number of data types that need to be specified precisely, with implementations that conform to specifications, is far too large for any typical R&D team to address. I.e., one needs to mobilize some community. In fact, even with many teams working towards the “right” shared vision[[1]](#footnote-1), accomplishing something approaching a “complete” solution will take a decade or longer.

An essential (but not sufficient) tool for this is a sharable, readily understood and implemented framework upon which contributions from (relatively) independent teams can be built and shared. This is commonly called an extensible information platform.

The Java programming language and Java environment have a near ideal structure for this.

#### What Existing Regli Code Can be Exploited?

Digital Archiving and Retrieval Tool (DART)12. DART fulfills the OAIS role of Ingest, i.e., it accepts files from a PSL-defined workflow, packages them, saving the context and metadata of the files, and adds the package into an archive. DART has a client and a server piece, the client is integrated with an engineer's desktop CAD work environment (e.g., Pro/ENGINEER) and has three high level goals: (1) Changes to the engineers' workflow must be minimized, or the archival system will not be utilized; (2) Capture as much meaningful data and context information as is practical, and at regular and key event intervals; (3) Automate the task of consistency checking, model verification and translation, i.e., when an archive snapshot is taken, in addition to capturing the native files, automatically translate model data into standard formats (e.g., STEP), visualization formats (e.g., mesh or 3D PDF). <http://gicl.cs.drexel.edu/wild/Digital_Archiving_and_RetrievaLTool> [[Regli 2009](file:///C:\\Users\\gladney\\AppData\\Local\\DigPres%20My%20Work\\TDO%20Design%20and%20Build\\Engineering\\Regli%202010%20Long-Term%20Retention%20of%20Geometriy-Centric%20Digital%20Engineering%20Artifacts%20(annotated).doc" \l "DART)]

### Background with Other Peoples‘ Work

#### HP & iPedigree

[HP and African Social Enterprise mPedigree Network Fight Counterfeit Drugs in Africa](http://www.hp.com/hpinfo/newsroom/press/2010/101206b.html) Free consumer text messaging service targets counterfeit pharmaceutical market that causes at least 700,000 deaths per year globally. Enables user verification of the authenticity of malaria medication and detect counterfeit drugs. <http://www8.hp.com/us/en/hp-information/social-innovation/mPedigree.html>

# Requirements

OEM requirements analysis: from [S.F. Personal Archiving CfP](Design%20and%20Build%20To%20Do.doc#Challenge),

### End-Users’ Specifications[[2]](#footnote-2)

What requirements would an LDP solution address? What might one of our descendants expect of information stored today? He would be satisfied if, for whatever record interested him, he could:

Retrieve a copy of the bit-string that represents the content if he is authorized to do so;

Read or otherwise use the content as its writers intended, without adverse effects caused by mistakes and inappropriate changes made by third parties;

Decide whether the information received is sufficiently trustworthy for his application;

Exploit embedded references reliably to identify and retrieve contextual information and to validate the trustworthiness of contextual links, doing so recursively to as much depth as he wants; and

Exercise all this functionality without hindrance by technical complexity that could be hidden.

In addition to professional authors, editors, archivists, and businessmen, some citizens will want to preserve information without needing to ask anybody’s permission to do so. They will want convenient tools and infrastructure to:

Package any content to be LDP-ready, doing so in some way that ensures that their descendants can use this content as specified immediately above;

Submit such readied content to repositories that promise to save it reliably, possibly in return for a fee for archiving service. (People are willing, in anticipation of death, to pay for storing their body remains. Surely they can be persuaded to pay for storing their intellectual remains together with high quality provenance information.)

What technology will repository institutions want? In addition to perfect world digital library technology, they will want support for:

Continuing to use their currently deployed repository software without disruption originating in extensions for LDP; replacing parts of this software in future years to provide their clients with the best services, doing so without disturbing already preserved information;

Sharing preserved content and metadata without adjustments requiring human judgment;

Sharing preservation effort with their clients to avoid burdens beyond repository resources; and

Ensuring that preserved information survives the demise of a large subset of all repositories.

Human users will want every step to be as automated as it can be without interfering with their subjective choices. Eliminating the distraction of clerical tasks will free them to focus on those authoring, editing, selection, and preservation activities that only human beings can accomplish.

### Software Reusability and Information Portability[[3]](#footnote-3)

The following requirements are additions to whatever is implied by TDO architecture articulated in ***PDI***.

1. Any PDO should be importable into any widely used digital repository offering.
2. LTDP extensions required for repository software should be limited to methods for extraction PDO content in order to construct digital library catalogs supporting search and administrative information. They should not alter how any repository operates other than by enabling it to handle PDO content.
3. LTDP extensions required for repository software should be provided in the form of a PDO toolkit that most repository administrators can easily interface to their installed repository implementations.
4. The PDO implementation should permit any genre of content whatsoever, requiring at most additions to handle each new data format wanted. Adding support for a new data format should not adversely affect the support for any already-enabled format.
5. A human user who wants to extract for editing (or for any other application) a content object from a PDO should be able to use his usual tools for that information format.
6. A human user who wants to extend a PDO by adding content, by updating content, or by adding a new representation of prior content should be able to do this within the rules that enable unsullied preservation of prior versions of the contained information.[[4]](#footnote-4)
7. Every content object[[5]](#footnote-5) should be provided with at least two identifiers:[[6]](#footnote-6) a UUID (***PDI §7.3.2***) that binds all instances of the information at hand, even instances containing changes relative to other instances, and a DRI (***PDI §7.3.4***) distinguishing version instances whose creator consider significant variants of the information at hand.[[7]](#footnote-7)

In contrast to the above requirements, other aspects are believed to be very much less important. In particular, computing resource costs needed to manage preserved objects is thought to be relatively unimportant because, at least for the foreseeable future, preserved information will be tiny compared to other information and will be inspected/manipulated only infrequently and only after some user is quite certain that certain preserved information is quite important to him/her. Even then, the principle use of PDOs will be to extract content that is then inspected/manipulated with pre-existing tools rather than our new software.[[8]](#footnote-8)

Address proprietary software, including the right to create derivative works.

### Comments on Metadata

The article “Managing Scientific Data” by Anastasia Ailamaki, Verena Kantere, and Debabrata Dash (June 2010) explained that data generated by research projects is valuable only when annotated with metadata describing the data’s provenance, context, and meaning. However, a given data item can be annotated in more ways than one, for two reasons:

***Provenance.*** A multidisciplinary project can track its progress with basic metadata indicating the provenance of its samples and their associated data. Each data item can also be annotated in a more detailed way through tools particular to the technique used to generate the data item; these annotations are themselves interpretable by people (and software) in the relevant discipline; and

***Assumptions.*** By definition, a research field involves a basic set of concepts used to understand the field but that is not yet agreed upon. Annotations beyond where and when the data was recorded incorporate assumptions that may be contentious among experts.

Data storage and metadata should thus be decoupled. A data repository must be capable of returning any data item stored within it, along with a list of places needed to find the relevant metadata. A metadata repository must be capable of identifying the schema it adheres to and respond to queries about specific data items with relevant annotations.

Decoupling the architecture this way eases develop of an ecosystem of repositories and annotation schemas. [Chris Morris, Warrington, U.K., CACM 53(9), 7](http://delivery.acm.org/10.1145/1820000/1810891/communications201009-dl.pdf?ip=99.8.229.65&acc=OPEN&CFID=52399848&CFTOKEN=54899685&__acm__=1320600563_d077523dc94b36555401420e15ed6875)

# Fundamentals from 20th-Century Epistemology

Conclusions extracted from ***PDI***.

The proper entities for archival attention are patterns inherent in transmitted and stored messages.[[9]](#footnote-9)

Digital repositories—sometimes called Content Management (CM) services—have been thoroughly understood for more than a decade, but include insufficient mechanism for reliably assuring information authenticity and intelligibility for decades or longer. We can mitigate the residual risks by adding to widely deployed CM offerings.

# Design

## High Level

Programming will be done with Java to ensure portability across platforms. Potential performance issues consequent on this choice will not be addressed until it is clear that they are real issues and not until the functionality and portability of all solution components are relatively refined.

Will include three representations of a PDO:

1. For manipulation by end users, a Java DOM representation in tree form. This will be used to recover/create/manipulate the other forms, and will include GUI interfaces that are as convenient to relatively naïve users as we are capable of making them.
2. For transmission between users and/or computing platforms, a linear form of the same information.[[10]](#footnote-10)
3. For storage in (local) repositories and also for manipulation by pre-existing tools (e.g., MS Office software for documents that it is designed for) in files collected into directories,[[11]](#footnote-11)with the top level of a PDO represented by a directory that might contain sub-directories.[[12]](#footnote-12) If some user wants to manipulate some content object, it will be necessary for him/her to have acquired and installed the (type-dependent) supporting software.[[13]](#footnote-13)
4. For efficiency, an optional addition would be an OWL/RDF representation.[[14]](#footnote-14) (Cf. [[Regli 2010] Figure 6](file:///C:\Users\gladney\AppData\Local\DigPres%20My%20Work\TDO%20Design%20and%20Build\Engineering\Regli%202010%20Semantics%20for%20Digital%20Engineering%20Archives%20(annotated).doc#OWLandXML).)

[Regli 2010] “[ontologies](file:///C:\\Users\\gladney\\AppData\\Local\\DigPres%20My%20Work\\TDO%20Design%20and%20Build\\Engineering\\Regli%202010%20Semantics%20for%20Digital%20Engineering%20Archives%20(annotated).doc" \l "ontologies) can also be used to record the relationships of a file to different versions of itself as it evolves over time and can record provenance meta-data about a file such as the creating agent, time, and location.”

Partly to avoid size problems and partly because direct manipulation of content objects will not be needed, these should not be imported into the DOM representation, but instead merely represented by pointers to the storage representation. Of course their representations should be included in the stream representation.

### Parameter Management

#### Design Parameters (for SW Engineers)

Reserved names in relationship table:

* “parent of” relationship
* “editor for” data type

Name and description of editor for data type, incl. predecessor and O/S

#### Data Type Parameters Common for All Users

For each O/S, the name of application to edit each type of content data

### PDO Structure and Update Management

Layout in storage is independent of logical structure

Logical structure as a graph (more general than a tree) in which each content object has some number of parents (the whole object is a root of the structure). “parent of” relationship in relationships table.

[[KEEP 2009](Bibliography.docx#KEEP)]

### Relationship Table

Make it four columns, with one column for operating system? (null or empty if “don’t care”). Or make it five columns to represent **<x relationship#1 y AND relationship#2 z>**?

## Further Considerations

The only thing I need java DOM (or JDOM) for is navigation within and control of a local stored version.

Stored version can be a single directory, and that can include a file representation of the Protection Block.

Calculate and save a verification key of each content block independently. Then calculate a verification key of an array of verification keys and other metadata.

### Conversion Storage-to-Linear and Reverse

Can I use a zip method, which is likely to be as efficient and as portable as anything I code myself?

### Notes 25th October

Was aware of this PREMIS/OWL work, and will read it in due course.

A very cursory scan suggests that it is a possibly useful addition to metadata schema favored by the academic library community, and by nobody else as far as I know. 🡺 it might fit into the broadly applicable preservation packaging toolkit that I've been working on, but only in a late stage, as an option for the indicated community.

That will be the case only if they have chosen a highly portable implementation environment, or code that can be automatically converted to fit the language that I've chosen--and if their code is of high quality.[[15]](#footnote-15)  Specifically, all my new code[[16]](#footnote-16) will be Java (which is readily converted to/from XML) and it will carry both payload and metadata blobs in whatever form/format their originators choose.  (Virtues of the latter is that recipients will have access to such payload elements with the usual tools for these), that this will work with almost no effort on my part, and that, whenever replacement support for these becomes available, it will automatically be usable on preserved content.)

With one interfering "gotcha", I'm making good progress towards a preservation tool that I believe can be made to please anybody who feels an urge to save stuff for a while, or to share stuff currently in situations for which evidence of unimpeached authenticity is important, (e.g., legal document collections, engineering document collections, military stuff, medical patent records).  It is pertinent that, for every application suggested, any adequate information sharing will include a contextual "surround" for whatever document(s) are the immediately interesting payload, with all the implications that you might think of for including such a "surround".[[17]](#footnote-17)

Such applications have equal importance to, and far greater revenue potential than, the focal topics of librarians and professional archivists.

**See attachments to Mike’s note for things to grab.**

# Bibliography

|  |  |
| --- | --- |
|  |  |
|  |  |
| [Abrams 2003] | Abrams, Stephen L. and David Seaman, [*Towards a global digital format registry*](http://archive.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf)*,* IFLA 2003.  The concept of format permeates all technical areas of digital repository architecture and operation. Proper interpretation of otherwise opaque content streams is dependent upon knowledge of how typed content is represented. The international digital library and archival communities have expressed their need for a sustainable, global registry of digital representation formats for the purpose of fulfilling their mission. The current MIME Media Types registry does not provide sufficient granularity of format typing or sufficient standardized representation information about formats. We present a summary of an ongoing international effort to establish a new global format registry, which will maintain persistent, unambiguous bindings between public identifiers for digital representation formats and the significant syntactic and semantic properties of those formats. |
| [BagIt progress 2011 Oct] | RESTful Bag Server spec - as found on our github site: <https://github.com/acdha/restful-bag-server>  We discussed the importance at this point of heading in the direction of furthering some of the initial implementations of the spec, and encouraging any other interested groups and parties to begin building out some preliminary implementations. A couple of implementations already underway include:  Play Framework Implementation: <http://groups.google.com/group/digital-curation/browse_thread/thread/0e8dfdb8c5a3e2aa?pli=1>  Java Bag Server:  [http://groups.google.com/group/digital-curation/browse\_thread/thread/60c070bdee101ef3#](http://groups.google.com/group/digital-curation/browse_thread/thread/60c070bdee101ef3) |
| [Bekaert 2003] | Bekaert, Jeroen, Patrick Hochstenbach, and Herbert Van de Sompel. *Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library*. D-Lib Magazine 9 (November 2003). |
| [Bekaert 2005] | Bekaert, Jeroen and Herbert Van de Sompel. *A Standards-based Solution for the Accurate Transfer of Digital Assets.* D-Lib Magazine 11 (June 2005). |
| [Bootz 2009] | Philippe Bootz, Samuel Szoniecky and Abderrahim Bargaoui, 2009. [*Entity/Identity: A tool designed to index documents about digital poetry*](http://archivesic.ccsd.cnrs.fr/docs/00/60/35/88/PDF/entity_identity.pdf), Symposium e-poetry, Barcelone, 2009.  L'article montre l'impact du phénomène de labilité sur les productions numériques, impact qui rend inopérantes les solutions de préservations utilisées par les musées pour préserver les installations numériques. Il propose alors une alternative fondée sur un modèle ontologique de l'oeuvre numérique qui sur la conception ontologique de Spinoza interprétée par Deleuze. L'article montre que cette ontologie est compatible avec le dispositif de communication créé par l'oeuvre. Ce modèle aboutit à une définition de l'oeuvre comme pouvoir d'agir. Ce modèle est enfin utilisé pour concevoir un outil d'indexation des oeuvres. |
| [Bootz 2011] | [*Towards an ontology of the field of digital poetry*](http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CCIQFjAB&url=http%3A%2F%2Felmcip.net%2Fsites%2Fdefault%2Ffiles%2Fattachments%2Fcriticalwriting%2Fbootz_presentation.ppt&ei=moOwTtaZLsWIsQKc3cDjAQ&usg=AFQjCNGSkbFOjF38Ce8y6K8_gLXt0gOQyg) elmcip.net/sites/default/files/attachments/.../**bootz**\_presentation.ppt  File Format: Microsoft Powerpoint - [Quick View](https://docs.google.com/viewer?a=v&q=cache:aX9Zn4rN-8QJ:elmcip.net/sites/default/files/attachments/criticalwriting/bootz_presentation.ppt+&hl=en&gl=us&pid=bl&srcid=ADGEESgo9rwoI1q17etWd5ccSGLADrxOU71PKNm0AyskZjb-59gnsmF-M5c-a9pA5-bM7iksVVadfM3-FB3mvSI6-udhG7Co3iMqIXA_P4ojU7vnGZj_hpZuQt7wOYumR2DSd5GQVfej&sig=AHIEtbQjY2O6I6I-_PWHBMIsNRiFTxZBRQ) Philippe Bootz & Samuel Szoniecky,. Laboratoire Paragraphe. Université Paris 8. 2. Entities/association model. 1-n and 0-n are cardinalities of the association |
| [Chan 1998] | [Patrick Chan](http://www.amazon.com/s/ref=ntt_athr_dp_sr_1?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Patrick%20Chan), [Doug Kramer](http://www.amazon.com/s/ref=ntt_athr_dp_sr_2?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Doug%20Kramer), [Rosanna Lee](http://www.amazon.com/s/ref=ntt_athr_dp_sr_3?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Rosanna%20Lee), *The Java Class Libraries, Volume 1: java.io, java.lang, java.math, java.net, java.text, java.util* ISBN 0201310023  This definitive reference to the Java 1.1 class libraries from the creators of the Java technology at Sun Microsystems is an essential resource for both novice and experienced Java programmers. The packages covered in this book are: java io, java net, java lang, java text, java lang.reflect, java util, java math, java util.zip. The 24,000 lines of code in over 600 examples facilitate learning-by-example and provide useful code fragments  [Patrick Chan](http://www.amazon.com/s/ref=ntt_athr_dp_sr_1?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Patrick%20Chan), [Rosanna Lee](http://www.amazon.com/s/ref=ntt_athr_dp_sr_3?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Rosanna%20Lee), The Java™ Class Libraries, Volume 2: java.applet, java.awt, java.beans (2nd Edition) ISBN 0201310031  This volume provides comprehensive reference documentation for the development of applets, user interfaces, and Java beans. The packages covered in Volume 2 are: java applet, java awt image, java awt, java awt peer, java awt data transfer, java beans, java awt event.  [Patrick Chan](http://www.amazon.com/s/ref=ntt_athr_dp_sr_1?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Patrick%20Chan), [Doug Kramer](http://www.amazon.com/s/ref=ntt_athr_dp_sr_2?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Doug%20Kramer), [Rosanna Lee](http://www.amazon.com/s/ref=ntt_athr_dp_sr_3?_encoding=UTF8&sort=relevancerank&search-alias=books&field-author=Rosanna%20Lee), *The Java(TM) Class Libraries: Supplement for the Java(TM) 2 Platform, v1.2; Parts A and B (Volume 1, Standard Edition)*  Contains descriptions of the new pacakges and classes introduced in the Javao 2 Platform, Standard Edition, version 1.2. It also provides details on how the classes and methods have changed from JDKo 1.1. Together, this supplement and The Java Class Libraries, Second Edition, Volume 1, provide a complete reference to the following packages: java.io, java.lang, java.lang.ref, java.lang.reflect, java.math, java.net, java.text, java.util, java.util.jar, and java.util.zip. The extensive class and member descriptions contain details crucial for developing robust and professional applets and applications. Each description is supplemented by an example that demonstrates the class or member in a relevant context. Numerous examples facilitate learning-by-example and provide useful code for your projects. The information is presented in an easy-to-use, dictionary-like format.Each class description includes: \*a class hierarchy diagram showing its connection to related classes \*a detailed overview describing its purpose and key concepts \*a convenient member summary that briefly describes each member and arranges the members into related groups \*an example demonstrating the class in a "real-world" context \*comprehensive descriptions and an example for each member |
| [D’Anjou 2005] | D’Anjou, Jim, Scott Fairbrother, Dan Kehn, John Kellerman, and Pat McCarthy. *The Java Developer’s Guide to Eclipse*, Addison-Wesley, 2005, ISBN 0-321-30502-7. Includes a CD.  Eclipse is an open source integration platform that contains a Java development programming environment. In other words, it can work seamlessly with other application development tools you need to use to get your job done. This in-depth and comprehensive book follows and supports the Eclipse user's learning curve. This book begins with the ins-and-outs of Eclipse and teaches the reader how to become a power user and use the flexibility of Eclipse to maximize their productivity. This comprehensive reference is also very practical and shows readers how to build their own plugins and create their own user interfaces. Like the first edition, this revision includes extensive exercises that reinforce all of the concepts from the book. This edition is updated to reflect all major changes in Eclipse 3.0 which include major changes to the APIs, plug-ins, the UI, widgets, JDT (Java development tools) and the ability to be used as a rich-client platform. The authors' extensive experience with training and teaching Eclipse to vendors makes this title a must-have for any Java developer learning Eclipse. |
| [Dunkley 2010] | Dunckley, Matt. Shahar Ronen, Ealan A. Henis, Simona Rabinovici-Cohen, Petra Reshef, David Giaretta, Esther Conway, *Using XFDU for CASPAR Information Packaging*, OCLC Systems & Services: International Digital Library Perspectives, Vol. 26 No. 2, 2010. |
| [Flanagan 2004] | Flanagan, David. 2004. *Java Examples in a Nutshell* ISBN 0596006209  Real-world Java programming examples that you can learn from. If you learn best "by example," this is the book for you. This third edition covers Java 1.4 and contains 193 complete, practical examples: over 21,900 lines of densely commented, professionally written Java code, covering 20 distinct client-side and server-side APIs. It includes new chapters on the Java Sound API and the New I/O API. The chapters on XML and servlets have been rewritten to cover the latest versions of the specifications and to demonstrate best practices for Java 1.4. New and updated examples throughout the book demonstrate many other new Java features and APIs. *Java Examples in a Nutshell* is a companion volume to *Java in a Nutshell*, *Java Foundation Classes in a Nutshell*, and *Java Enterprise in a Nutshell*. It picks up where those quick references leave off, providing a wealth of examples for both novices and experts. This book doesn't hold your hand; it simply delivers well-commented working examples with succinct explanations to help you learn and explore Java and its APIs. *Java Examples in a Nutshell* contains examples that demonstrate:   * Core APIs, including I/O, New I/O, threads, networking, security, serialization, and reflection * Desktop APIs, highlighting Swing GUIs, Java 2D graphics, preferences, printing, drag-and-drop, JavaBeans, applets, and sound * Enterprise APIs, including JDBC (database access), JAXP (XML parsing and transformation), Servlets 2.4, JSP 2.0 (JavaServer Pages), and RMI   The book begins with introductory examples demonstrating structured and object-oriented programming techniques for new Java programmers. A special index at the end of the book makes it easy to look up examples that use a particular Java class or accomplish a desired task. In between, each chapter includes exercises that challenge readers and suggest further avenues for exploration. |
| [Flanagan 2005] | Flanagan, David. 2005. *Java in a Nutshell* ISBN 0596007736  With more than 700,000 copies sold to date, *Java in a Nutshell* from O'Reilly is clearly the favorite resource amongst the legion of developers and programmers using Java technology. And now, with the release of the 5.0 version of Java, O'Reilly has given the book that defined the "in a Nutshell" category another impressive tune-up.  In this latest revision, readers will find *Java in a Nutshell*, 5th Edition, does more than just cover the extensive changes implicit in 5.0, the newest version of Java. It's undergone a complete makeover--in scope, size, and type of coverage--in order to more closely meet the needs of the modern Java programmer.  To wit, *Java in a Nutshell*, 5th Edition now places less emphasis on coming to Java from C and C++, and adds more discussion on tools and frameworks. It also offers new code examples to illustrate the working of APIs, and, of course, extensive coverage of Java 5.0. But faithful readers take comfort: it still hasn't lost any of its core elements that made it such a classic to begin with.  This handy reference gets right to the heart of the program with an accelerated introduction to the Java programming language and its key APIs--ideal for developers wishing to start writing code right away. And, as was the case in previous editions, Java in a Nutshell, 5th Edition is once again chock-full of poignant tips, techniques, examples, and practical advice. For as long as Java has existed, *Java in a Nutshell* has helped developers maximize the capabilities of the program's newest versions. And this latest edition is no different. |
| [GFDR] | [Global Digital Format Registry (GDFR) Information Site](http://www.gdfr.info/) [www.gdfr.info/](http://www.gdfr.info/) The Global Digital Format Registry! (GDFR) is meant to be a distributed and replicated registry of format information populated and vetted by **...** |
| [Giaretta 2011] | Giaretta, David. [*Advanced Digital Preservation*](http://www.amazon.com/gp/product/3642168086), 2011, ISBN 3-642-16808-6.  Why is there a need for yet one more book on digital preservation? Because, for the most part, other publications focus on documents, images and webpages—objects that are normally rendered to be simply displayed by software to a human viewer. There are clearly many more types of digital objects that may need to be preserved, such as databases, scientific data and software itself. This book explains why the tools and techniques used for preserving rendered objects are inadequate for all these other types of digital objects, and provides the concepts, techniques and tools that are needed.  It is structured in three parts. The first part is on theory—concepts and techniques that are essential for preserving digitally encoded information. The second shows practice—use and validation of these techniques. The third addresses judging whether money is being well spent.  Examples of digital objects from many sources illustrate the tools and techniques. To assist readers, the book is supported by many hours of videos and presentations from the CASPAR project and by a set of open source software. |
| [Gladney 2007a] | Gladney, H.M. [*Digital Preservation in a National Context: Questions and Views of an NDIIPP Outsider*](http://www.dlib.org/dlib/january07/gladney/01gladney.html), D-Lib Magazine 13(1/2), January 2007.  This article draws attention to technical opportunities which, if pursued, would significantly accelerate National Digital Information Infrastructure Preservation Program (NDIIPP) progress towards objectives called for by the U.S. Congress. It also identifies concerns about apparent content scope limitations of the NDIIPP plan.  A solution is known in principle for every difficult technical problem of digital preservation, including all those identified in NDIIPP publications. They and other works correctly assert that non-technical preservation challenges are greater than technical ones, but do not discuss using technology to reduce non-technical obstacles. Available technical choices show that some apparent preservation challenges are not obstacles after all.  If document representations and network protocols are standardized, then each archive can autonomously adapt itself to its own institutional environment. Thinking about what end users will want led my colleagues and me to approach the challenge differently than most other authors. This article focuses on information contributors and readers instead of on the work of repository employees. It addresses the design of document representations instead of new repository methodology. Each repository is treated as a "black box" whose internals can be adapted to local needs instead of discussing sharable repository implementations. |
| [Gladney 2007b] | Gladney, H.M. *Economics and Engineering for Preserving Digital Content*, draft paper, 3-Dec-07. Available at <http://eprints.erpanet.org/139/01/LDP_Engineering.pdf>. [***My stored draft copy***](file:///G:\W\DigPres\DigPres%20My%20Papers\Background\LDP%20Engineering%20(submit).doc). |
| [Gladney 2009] | Gladney, H.M. 2009. *Long-term Preservation of Digital Records: Trustworthy Digital Objects.* The American Archivist 72(2), 401-435. ISSN: 03609081 [***copy***](file:///C:\Users\gladney\AppData\Local\DigPres%20My%20Papers\AmArchivist\Final%20draft%20(adapted%20with%20minor%20updates).doc)  The proper entities for archival attention are patterns inherent in transmitted and stored messages. Most digital archival repository technology that private sector enterprises call content management (CM) technology has been thoroughly understood and widely deployed for more than a decade. This technology is not adequate for long-term digital preservation because it includes no mechanisms for reliably assuring authenticity and intelligibility of digital documents for fifty years or longer. CM provides for near-term preservation without handling long-term preservation, which must overcome risks associated with technological obsolescence and fading human memory. This article offers a solution to mitigate these risks. Implementing the needed software would be a small addition to widely deployed CM offerings. This long-term preservation solution, devised for cultural and scholarly digital documents, is already structured to support archival principles for business records, and the article describes this Trustworthy Digital Object (TDO) architecture and its design core sufficiently to show how archivists can participate in managing digital repositories that conform and are attuned to the particular needs of any archival institution. |
| [Gladney 2011b] | Gladney, H.M. 2011. *Preparing Digital Information for Long-Term Preservation,* US Patent Application13219630, August 2011.  Describes a method for creating or extending a digital document represented so as to be durably intelligible and reliably trustworthy. Such a document embeds standardized metadata, provenance information, and reliable links to chosen documents within a world-wide network of digital repositories. These links and the documents’ own identifier(s) are chosen to uniquely, unambiguously, and forever identify what they refer to. This system provides a robustly durable method of preserving an unbounded number of digital objects for as long as their representing bit-strings are kept in existence and findable by now-conventional digital library technology, as first described in [Gladney 2000] and publications by the same author. The overall system provides this service without requiring that pre-existing software be modified, and without requiring that any information object that it is intended to protect be modified from what its authors provided. |
| [Gladney 2011c] | Gladney, H.M. 2011. *Maintaining Preserved Information,* US Patent Application13225520, September 2011.  Describes a method for updating preserved data objects to take advantage of software tools that have been improved since such information objects were created.  A preserved digital document includes within itself standardized metadata, provenance information, and reliable links to other documents stored within a world-wide network of digital repositories. It makes use of cryptographic sealing and a network of key-sharing sites to ensure information authenticity. This invention provides a robustly durable method of preserving an unbounded number of digital objects for as long as their representing bit-strings are kept in existence and findable by now-conventional digital library technology. The overall system provides this service without requiring that pre-existing software be modified, and without requiring that any protected information be modified from what its originators provided. |
| [Grüninger 2003] | M. Grüninger and C. Menzel*. The* [*process specification language (PSL) theory and applica­tions*](http://philebus.tamu.edu/cmenzel/Papers/Gruninger&Menzel-PSL.pdf)*.* AI Mag.,24(3):63-74, 2003. [***saved***](file:///C:\Users\gladney\AppData\Book2\Takings\Computing\Preservation\Gruninger%20Menzel%202003%20PSL%20process%20specification%20language.pdf)  PSL has been designed to facilitate correct and complete exchange of process information among manufacturing systems, such as scheduling, process modeling, process planning, production planning, simulation, project management, work flow, and business-process reengineering. We give an overview of the theories within the PSL ontology, discuss some of the design principles for the ontology, and finish with examples of process specifications that are based on the ontology.  PSL is standardized as [ISO 18629-14:2006](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35186). |
| [ISO 2001] | ISO 2001. *Automation systems and integration — Product data representation and exchange,* described at <http://en.wikipedia.org/wiki/ISO_10303-21>.  The format of a STEP-File is defined in ISO 10303-21 *Clear Text Encoding of the Exchange Structure*. ISO 10303-21 defines the encoding mechanism on how to represent data according to a given [EXPRESS schema](http://en.wikipedia.org/wiki/ISO_10303-11), but not the EXPRESS schema itself. STEP-File are also called *p21-File* and *STEP Physical File*. The file extensions *.stp* and *.step* indicate that the file contain data conforming to STEP Application Protocols while the extension *.p21* should be used for all other purposes.[[](http://en.wikipedia.org/wiki/ISO_10303-21#cite_note-1) |
| [McHenry 2009] | K. McHenry and P. Bajcsy. *Key aspects in 3D file format conversions*. In Joint Annual Meeting of the Society of American Archivists and the Council of State Archivists, August 11, Hilton Austin, Texas, USA 2009. [slides](http://www2.archivists.org/sites/all/files/PeterBajcsy-SAA-ResearchForum2009.pdf) [saved](file:///C:\Users\gladney\AppData\Book2\Takings\Computing\Preservation\McHenry%20Bajcsy%202009%20SAA%203D%20file%20format%20conversions.pdf) |
| [McLaughlin, 2006] | McLaughlin, Brett et al.. 2006. *Java and XML*, ISBN 059610149X  This third edition of *Java and XML* covers all major Java XML processing libraries, including full coverage of the SAX, DOM, StAX, JDOM, and dom4j APIs as well as the latest version of the Java API for XML Processing (JAXP) and Java Architecture for XML Binding (JAXB). The chapters on web technology have been entirely rewritten to focus on the today's most relevant topics: syndicating content with RSS and creating Web 2.0 applications. You'll learn how to create, read, and modify RSS feeds for syndicated content and use XML to power the next generation of websites with Ajax and Adobe Flash. Topics include:  Basics of XML, including DTDs, namespaces, XML Schema, XPath, and Transformations  The SAX API, including all handlers, filters, and writers  The DOM API, including DOM Level 2, Level 3, and the DOM HTML module  The JDOM API, including the core and a look at XPath support  The StAX API, including StAX factories, producing documents and XMLPull  Data Binding with JAXB, using the new JAXB 2.0 annotations  Web syndication and podcasting with RSS  XML on the Presentation Layer, paying attention to Ajax and Flash applications |
| [Ram 2006] | S. Ram and J. Liu. *Understanding the Semantics of Data Provenance to Support Active Conceptual Modeling*. Springer, 2006. |
| [Regli 2010] | William C. Regli, Joseph B. Kopena, Michael Grauer, Timothy W. Simpson, Robert B. Stone, Kemper Lewis, Matt R. Bohm, David Wilkie, Martin Piecyk, Jordan Osecki, [*Semantics for Digital Engineering Archives Supporting Engineering Design Education*](file:///C:\Users\gladney\AppData\Local\Other%20People's%20Work\Archiving\NIST\Regli%202010%20Semantics%20for%20Digital%20Engineering%20Archives%20(annotated).doc)*,* AI Magazine 31(1), 37-50, 2010. [***Annotated copy***](file:///C:\Users\gladney\AppData\Local\Other%20People's%20Work\Archiving\NIST\Regli%202010%20Long-Term%20Retention%20of%20Geometriy-Centric%20Digital%20Engineering%20Artifacts%20(annotated).doc)  This article introduces the challenge of digital preservation in the area of engineering design and manufacturing and presents a methodology to apply knowledge representation and semantic techniques to develop Digital Engineering Archives. This work is part of an ongoing, multiuniversity, effort to create cyber infrastructure-based engineering repositories for undergraduates (CIBER-U) to support engineering design education. The technical approach is to use knowledge representation techniques to create formal models of engineering data elements, workﬂows and processes. With these formal engineering knowledge and processes can be captured and preserved with some guarantee of long-term interpretability. The article presents examples of how the techniques can be used to encode speciﬁc engineering information packages and workﬂows. These techniques are being integrated into a semantic wiki that supports the CIBER-U engineering education activities across nine universities and involving over 3500 students since 2006. |
| [[Regli](http://gicl.cs.drexel.edu/wiki/William_C._Regli) 2011] | Regli, William C. [Joseph B. Kopena](http://www.informatik.uni-trier.de/%7Eley/db/indices/a-tree/k/Kopena:Joseph_B=.html), [Michael Grauer](http://www.informatik.uni-trier.de/%7Eley/db/indices/a-tree/g/Grauer:Michael.html). *On the long-term retention of geometry-centric digital engineering artifacts*. [Computer-Aided Design 43](http://www.informatik.uni-trier.de/%7Eley/db/journals/cad/cad43.html#RegliKG11)(7): 820-837 (2011) [***saved and HMG annotated***](file:///C:\Users\gladney\AppData\Local\Other%20People's%20Work\Archiving\NIST\Regli%202010%20Long-Term%20Retention%20of%20Geometriy-Centric%20Digital%20Engineering%20Artifacts%20(annotated).doc)  This paper discusses the challenges of long-term preservation of digital geometric models and the engineering processes associated with them. For engineering, design, manufacturing, and physics-based simulation data this requires formats that are accessible potentially indefinitely into the future. One of the fundamental challenges is the development of digital geometry-centric engineering representations that are self describing and assured to be interpretable over the long lifespans required by archival applications. Additionally, future users may have needs that require other information, going beyond geometry, be also accessible to fully interpret the model. These problems are highly interdisciplinary and not exclusively algorithmic or technical. To provide context, the paper introduces a case study illustrating an overall portrait of the problem. Based on observations from this case study, we present a framework for enhancing the preservation of geometry-centric engineering knowledge. This framework is currently being used on a number of projects in engineering education. |
| [Stancic 2011] | Stancic, Hrvoje. Kresimir Pavlina, Arian Rajh, *Long-term Preservation Solution for Complex Digital Objects Preserved as Archival Information Packages in the Domain of Pharmaceutical Records*, eTELEMED 2011 : The Third International Conference on eHealth, Telemedicine, and Social Medicine, 2011. [**Saved copy**](../Other%20People's%20Work/Stancic%202011%20LTDP%20Pharma.docx)  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. |
| [Viktor 2010] | Viktor, H.L.; Bo Wang; Paquet, E.; Doyle, J. *Preserving object-relational databases for the next generations*, 2010, [Digital Information Management (ICDIM), Fifth International Conference](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5644872) [**copy**](file:///G:\W\DigPres\Other%20People's%20Work\Viktor%202010%20Object-Relational%20DBs.pdf)  Increasingly, resources are “born digital” and their associated formats are short-lived. Subsequently, the development of environments to preserve such digital content over the very long-term (50 years or more) has become a critical issue. To date, however, the preservation of data as contained in object-relational databases has been widely overlooked. Here, the task is inherently complicated by the nature of the data (relational as well as multimedia). Furthermore, the internal structure of the database and the associated applications need to be preserved as it evolves over time. This paper presents an environment to preserve object-relational databases over a very long period of time. We show that our environment is able to host and access multiple databases as they evolve over time. |

### [Links](file:///G:\\W\\DigPres\\WIP\\Bibliography.docx" \l "LINKS) & [Current Manuscripts](file:///G:\W\DigPres\WIP)

[***Annotated Am. Archivist paper***](file:///G:\W\DigPres\DigPres%20My%20Work\TDO%20Design%20and%20Build\Engineering\Gladney%202009%20Am%20Archivist%20final%20draft%20(used%20with%20minor%20updates%20and%20annotated%20in%202011).docx) [Tools Summary](To%20Do.docx#ToolsSummary)

#### [*Critiques of other work*](Critique%20of%20TDR%20and%20Other%20Work.docx)

# To Do?

[**Prepare a 2012 conference paper**](file:///G:\W\DigPres\WIP\Design%20Engineering%204%20Copenhagen.doc)

<http://longtermdigitalpreservation.com/consultants/henry-gladney.htm>

Digital curation group: <http://groups.google.com/group/digital-curation/>

[Bibliography - Henry Gladney](http://www.hgladney.com/bibliography.htm) [www.hgladney.com/bibliography.htm](http://www.hgladney.com/bibliography.htm) Sep 24, 2011 - “In case you haven't read them yet, I'd recommend taking a look at Henry Gladney's article “Long-Term Preservation of Digital Records: Trustworthy Digital

[Principles for digital preservation](http://dl.acm.org/citation.cfm?id=1113038) dl.acm.org/citation.cfm?id=1113038 by HM Gladney - 2006 - [Cited by 27](http://scholar.google.com/scholar?hl=en&lr=&cites=18271468207831739537&um=1&ie=UTF-8&ei=BCiwTpGSHKOLiAKxsdn3Dw&sa=X&oi=science_links&ct=sl-citedby&resnum=12&ved=0CJoBEM4CMAs) - [Related articles](http://scholar.google.com/scholar?hl=en&lr=&q=related:kWynVIdLkf0J:scholar.google.com/&um=1&ie=UTF-8&ei=BCiwTpGSHKOLiAKxsdn3Dw&sa=X&oi=science_links&ct=sl-related&resnum=12&ved=0CJsBEM8CMAs) Jul 16, 2011 - Duranti, L. The long-term preservation of the dynamic and interactive records of ... Henry M. Gladney, Trustworthy 100-year digital objects: Evidence after every ...

[Digital Curation and Preservation Bibliography: 1 General Works](http://digital-scholarship.org/dcpb/general.htm) *digital-scholarship.org › ... ›* [*Table of Contents*](https://www.google.com/url?url=http://digital-scholarship.org/dcpb/toc.htm&rct=j&sa=X&ei=BCiwTpGSHKOLiAKxsdn3Dw&ved=0CKQBEOkFKAAwDA&q=Gladney+preservation&usg=AFQjCNHqu3FR_-TVogMJRXJ1cG9rmEFG2Q)Jan 29, 2011 - "Selection for Preservation in the Digital Age." Library Resources & Technical Services 44, no. 2 (2000): 97-104. Gladney, H. M. "Principles for Digital ...

[Digital Curation and Preservation Bibliography: 3 Formats and ...](http://digital-scholarship.org/dcpb/formats.htm) *digital-scholarship.org › ... ›* [*Table of Contents*](https://www.google.com/url?url=http://digital-scholarship.org/dcpb/toc.htm&rct=j&sa=X&ei=BCiwTpGSHKOLiAKxsdn3Dw&ved=0CK0BEOkFKAAwDQ&q=Gladney+preservation&usg=AFQjCNHqu3FR_-TVogMJRXJ1cG9rmEFG2Q)Jan 28, 2011 - Gladney, H. M., and R. A. Lorie. "Trustworthy 100-Year Digital ... "Formats for Digital Preservation: A Review of Alternatives and Issues." Information Services and ...

[Trustworthy 100-year digital objects](http://dl.acm.org/citation.cfm?id=1080346) dl.acm.org/citation.cfm?id=1080346 by HM Gladney - 2005 - [Cited by 34](http://scholar.google.com/scholar?hl=en&lr=&cites=11275725974747200746&um=1&ie=UTF-8&ei=BCiwTpGSHKOLiAKxsdn3Dw&sa=X&oi=science_links&ct=sl-citedby&resnum=36&ved=0CNoCEM4CMCM) - [Related articles](http://scholar.google.com/scholar?hl=en&lr=&q=related:6rxzXQhte5wJ:scholar.google.com/&um=1&ie=UTF-8&ei=BCiwTpGSHKOLiAKxsdn3Dw&sa=X&oi=science_links&ct=sl-related&resnum=36&ved=0CNsCEM8CMCM) Apr 10, 2011 - Gladney, H. M. 2005. Principles for digital preservation. Commun. ACM. To be published. Draft available online at http://eprints.erpanet.org/archive/00000070/. ...

[Preserving object-relational databases for the next generations](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5664663) *ieeexplore.ieee.org › ... ›* [*Conferences*](https://www.google.com/url?url=http://ieeexplore.ieee.org/xpl/conferences.jsp&rct=j&sa=X&ei=BCiwTpGSHKOLiAKxsdn3Dw&ved=0CJkDEOkFKAAwKg&q=Gladney+preservation&usg=AFQjCNHUs48FPtQjZ5hFxHKgG-1ldMu11Q) *›* [*Digital Information Management*](https://www.google.com/url?url=http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp%3Fpunumber%3D5644872&rct=j&sa=X&ei=BCiwTpGSHKOLiAKxsdn3Dw&ved=0CJoDEOkFKAEwKg&q=Gladney+preservation&usg=AFQjCNGfUnB0ljejnpH05J9RDdpDZlo8xw)by HL Viktor - [Related articles](http://scholar.google.com/scholar?hl=en&lr=&q=related:rljcL3GyTKwJ:scholar.google.com/&um=1&ie=UTF-8&ei=BCiwTpGSHKOLiAKxsdn3Dw&sa=X&oi=science_links&ct=sl-related&resnum=43&ved=0CJYDEM8CMCo)  
Dec 10, 2010 - Doyle, J., Viktor, H.L., Paquet, E.:., Preservation Metadata: A Framework for 3D Data based on the Semantic Web; Gladney, H. M., Lorie., R:, Trustworthy ...

# Candidate Figures

|  |
| --- |
| Flow overview (PDI Fig. 2) better.jpg |
| Figure 1: Human and machine roles in sharing documents (*PDI Fig. 2*) |

|  |
| --- |
| Flow simple.jpg |
| Figure 2: Human and machine roles in sharing documents (simplified version) |

|  |
| --- |
| Patent repositories model (PDI Fig.32).jpg |
| Figure 3: Model of digital repositories and related human roles (*PDI Fig. 26*) |

|  |
| --- |
| TDO (detailed).jpg Patent TDO.jpg |
| Figure 4: Schema for a Trustworthy Digital Object (part of *PDI Fig. 32*) |

|  |
| --- |
| PIP 2011 Sept.jpg |
| Figure 5: Schema for a Preserved Information Package |

|  |
| --- |
| Protection Block.jpg |
| Figure 6: Content of a TDO Protection Block (*PDI Fig. 33*) |

# End Document

1. Although this is not specified here, I believe it can readily be specified in a form that will, over time, win enthusiastic participation. [↑](#footnote-ref-1)
2. ` From [Gladney 2009 §What Would an LDP Solution Accomplish]. [↑](#footnote-ref-2)
3. These technical requirements overlap the end-users’ requirements somewhat, but are worded to be more clearly understood by software engineers. [↑](#footnote-ref-3)
4. This alludes to TDO support for nesting prior TDOs within new TDO versions, and also to the linking mechanism for new content representations to their predecessors. [↑](#footnote-ref-4)
5. I.e., this rule applies both to PDOs, also to content objects held within PDOs, and also to related information objects that might not (yet) be held within any PDO. [↑](#footnote-ref-5)
6. See the discussion of identifiers in ***PDI §7.3****.* [↑](#footnote-ref-6)
7. The relationship of an object identifier to the object(s) it identifies is a subjective choice rather than an objective fact. See ***PDI §7.3.1.*** [↑](#footnote-ref-7)
8. For instance, a PDO might convey what was originally an MS Word document. To inspect and otherwise manipulate it, a user will employ (a future version of) MS Word after using PDO software to create a Word-accessible copy of it. [↑](#footnote-ref-8)
9. Our earlier solution, devised for preserving cultural digital documents, accepts metadata implementing archival principles for business records. [↑](#footnote-ref-9)
10. Java DOM support includes tools for conversion between tree and linear forms. [↑](#footnote-ref-10)
11. Obviously, the details of the file/directory structure will be operating-system dependent. [↑](#footnote-ref-11)
12. We might want to use “.zip” and/or “.jar” support for long-term storage. (TBD) [↑](#footnote-ref-12)
13. E.g., to edit a word processor object, the user will need an appropriate word-processing editor. (We cannot manage this for users, partly for licensing reasons, partly because we cannot anticipate what data types are carried in PDOs, and partly because software versions will evolve with time.) [↑](#footnote-ref-13)
14. There exist tools for DOM-to-OWL/RDF conversion. See [***notes***](file:///C:\Users\gladney\AppData\Local\Temp\Design.docx#JavaAndOWL). [↑](#footnote-ref-14)
15. E.g., any Microsoft Office file will be accessible via the appropriate MS Office component. [↑](#footnote-ref-15)
16. For the time being at least, I will refer to what I am implementing as "PIP" ("preserved information package") to differentiate it from "TDO", because the architecture might diverge as a consequence of discoveries made during implementation.  This choice does not imply anything about naming of what is eventually released. [↑](#footnote-ref-16)
17. In this, the notion of "complete" context is almost meaningless.  What's enough context will always be a subjective decision. [↑](#footnote-ref-17)