# Long Term Management of Private Digital Assets

Heinrich Jasper
TU Bergakademie Freiberg, Germany
jasper@informtik.tu-freiberg.de

## Abstract

The lifetime of each individual is accompanied by a growing amount of digital assets, e.g. documents, pictures etc. Important digital assets will exist longer than the typical lifespan of the respective technology used for its creation, i.e. for several decades. We are looking for supportive technologies that help individuals in preserving their digital assets. Furthermore, documents might be involved in transactions with a third party, a topic that has to be handled for long-living digital assets, too.

Our work concentrates on basic technologies that treat these digital assets as first class citizens in a changing digital environment called cyberworld in the sequel. Most important for us are ontologies and active agents based on a reliable data management infrastructure. First results provide helpful insight in the next steps to be taken along the road to a secure, trustfully environment that helps individuals in their long term management of digital assets.

## 1. Introduction

We focus in our paper on long living digital assets that accompany an individual during a long period of his or here lifetime and even sometimes beyond that. Consider a typical knowledge worker in the western world who registers after school at some university in order to get an academic degree. Thereafter, this individual intends either an academic career or wants to apply for a job in industry or at a public authority.

At the same time, this individual's progress is accompanied by a lot of private documents, e.g. reports, references, certificates as well as insurance contracts and so on. These documents will be all digital in the near future, for instance due to e-government activities in the European Union or as part the e-commerce activities of insurance companies. Furthermore, each individual will gather a lot more digital assets because of more or less intensive usage of state-of-the-art digital gadgets like cameras, personal digital communicators and so on. Consider for example digital photos taken at the graduation ceremony that are intended to be shown to the individual's grandchildren.

Whereas nowadays important assets like legal documents are preserved on paper, our intended future scenarios will see all these assets being available digitally and exchanged electronically within a state-of-the-art networked environment, e.g. via the internet. Consider the admission process for a university as an example. The following key elements must be handled in an all-digital admission:

- The admission is done completely on-line.
- The individual who wants to join the university – the candidate – selects the admission process from the universities internet portal.
- The candidate fills out all forms and signs them.
- The candidate adds all necessary documents, i.e. the school report and eventually some additional references.
- The admission process is carried out by the universities personnel:
  a) if successful, the individual gets some legal document of his student status.
  b) If not successful, the candidate is informed and the data are stored at the university

Figure 1 visualizes the administration process partially.

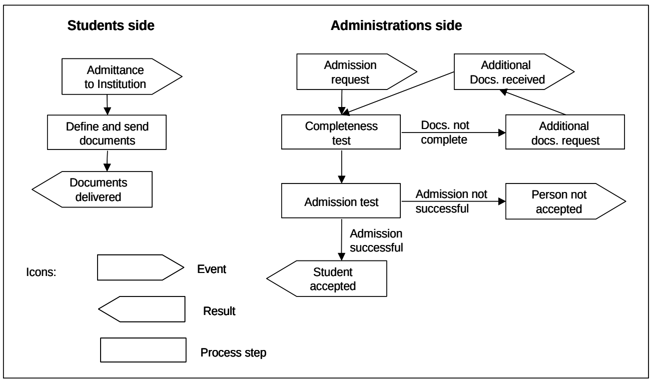


Figure 1. Admission processes (part.)

Besides the implementation of such all-digital real world processes we are concerned about the management of the digital assets that are involved in suchlike processes. These digital assets are characterized by special properties:

1. *Longevity*: each digital asset has to exist and to be accessible by an individual much longer than the typical lifespan of the technical infrastructure it was created in and
2. *Ownership*: each digital asset is controlled completely by its owner, i.e. by its creator or by a third party or by the legal successors of one of the former.

A third property is assigned to a subset of these digital assets. Elements of this subset are used in formal transactions in order exchange information with a third party, for example in order to establish a contract or in the admission process exemplified above.

3. *Transactionality*: such a digital asset has a value in a negotiation or treaty with a third party.

Examples of this kind of digital asset are documents, e.g. the diploma of a higher education institution that admits the individual for university studies. The individual will have at least access to it for her or his lifespan.

A long living private digital asset is called LPDA in the sequel. LPDAs having all three properties are in the focus of our research and studied in the reset of this paper.

## 2. Requirements

There exists a whole bunch of requirements for the long term secure and trustful management of these LPDAs. We concentrate on the following requirements in our work:

### 2.1. Continuous Accessibility and Control

Each document must be accessible during its predictable lifetime. The lifetime for a lot of documents is connected to the lifetime of their owner or even lasts longer. Accessible means both search- and readability for humans by natural senses as well as exchangeability with third party infrastructures.

Since we extrapolate changing hardware and software infrastructures in the foreseeable future this requirement results in the problems due to long term digital data preservation, see for example [1]. Besides a large set of approaches all over the world (see for example in [2] the NESTOR project by the BMBF the German Ministery of Education and Research) to tackle this problem there is no satisfying solution known so far. In addition to archaeological methods (i. e. conserving the complete hard- and software infrastructure together with the digital assets) migration and emulation techniques are under consideration, see e. g. [3].

Since we are considering commonplace usage of assets by everyone the preservation requirement can only be solved by providing means for accessibility of the digital assets in the state-of-the-art infrastructure of each individual. Due to the extrapolated changing of hardware and software infrastructures this results in establishing a migration process that is mostly transparent to the user in addition to the traditional security procedures for the management of data like backups etc.

### 2.2. Long Term Security and Exchangeability

In order to establish private-public processes of exchanging digital assets (e. g. legal documents, see the above mentioned scenario of admission to university as an example) in the long run several requirements have to be fulfilled.

Digital assets used in exchange processes must always be in the same state as the “original” digital asset. This requires mechanisms for guaranteeing that assets are unaltered to prevent no-repudiation by the third party. State-of-the-art techniques are electronic signatures based on public key infrastructures (PKI). Each legal document must be signed by its originator and there has to be an infrastructure for defining, accepting and exchanging legal documents. Since the signature provided by the PKI expires at some predefined point of time in the future, additional means must be established to re-signature the document. This should be transparent to the user or owner of the document.

The public key infrastructure results in additional requirements. Each individual involved in the above mentioned processes has to be identified. There are several approaches known so far. Internet based single sign on solutions are one possible approach. Other approaches are identities provided by authorities, for instance the Germany smart card or biological access technologies.

To sum up, the questions to be solved from the individual’s point of view are:

- Where do I store my data safely, i.e. how can I be sure that my digital assets exist and are accessible in decades?
- How do I organize the storage of digital assets on my own and how do I find and use them for private-public processes when needed, e.g. in several years?
- Who is in charge for providing digital signatures (e. g. what trust centres will exist in Germany in the long run)?

These requirements are up to now solved to some extent by using a third party for handling private data, for instance a service centre of a bank consortium, see [4]. Nevertheless very few people use this scenario up to now. On the other hand, the management of private assets is in most cases done via standard personal computer infrastructure based on one of the two operating system families either Windows from Microsoft or Unix either by Apple or the Linux communities. Although the approach based on own infrastructure is common for managing most of the individual's private digital assets, the questions above are unsolved so far.

## 3. Private Digital Asset Management

We root our approach for the management of LPDAs on a (mostly) transparent migration technology. Thus any asset stored in the individual's digital infrastructure (IDI) will remain there in the foreseeable future. The IDI is an assumption on properties of the hard- and software infrastructure that will be detailed later. The basic idea is a continuous migration of each asset. Each individual's IDI changes more or less abruptly in specific points in time. Furthermore individual IDIs exist in a networked environment called cyberworld that has possibly effects on the LPDAs stored in the IDIs, see e. g. the outdating of electronic signatures mentioned above. Without loss of generality we assume the cyberworld as a network of IDIs.

### 3.1. Services for Digital Asset Management

In order to manage LPDAs each IDI must establish a set of functions that might (should) be transparent to the owner of the IDI, the so called LPDA protocol, which is a set of services. Besides as set of auxiliary functions the main functions to the LPDA protocol are (for a more complete list of functions look at [5]):

1. Visible functions for the user:
   a. *Insert Asset*: A new asset or a set of assets (e. g. digital photo series) is integrated into the PDAM system.
   b. *Retrieve Asset*: An asset (or a set of assets) described by a set of properties is selected and reissued in defined format.
   c. *Sign Asset*: An asset gets the electronic signature of its owner.
2. Internal – non-visible functions
   a. *Migration*: since manual migration is not applicable here because of the required non-visibility to the user there must be some sort of automatic migration. Three main approaches are known in the literature, cf. [6]:
      i. *OAIS-Standard*: The open archive information systems (OAIS-) standard defines necessary functionality but not the techniques. We get on with this approach as discussed later on.
      ii. *Virtual universal computer*: Based on a virtual computer architecture (cf. [1]) abstract methods and data schemata are used for storage and retrieval. An implementation of the virtual system is necessary only at access time.
      iii. *Preservation-Layer-Method*: This approach models the infrastructure explicitly and knows about the implementation of each function necessary for handling a digital asset.
   b. *Re-signaturing*: Re-signaturing is a critical function for trustful long term asset management. This function should be self-acting and thus non-visible to the user. There are two different cases here:
      i. *Own assets*: In this case the IDI of the user is able to re-signature the asset along the guiding principles of the PKI in use.
      ii. *Third party assets*: In order to re-signature such an asset the third party must provide means to bind a new signature to it; this is under complete control of the third party.

The functionality for long term digital asset management using IDIs is summarized in figure 2.

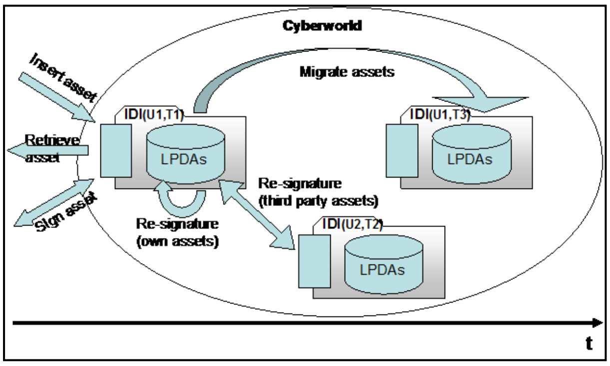


Figure 2. Digital Asset Management in Time

### 3.2. Architectural Issues

Our approach uses a combination of known technologies in order to empower IDIs with the desired functionality. These are:

1. *Virtual machine*: Each individual digital infrastructure provides functionality that is defined by a virtual machine for the management of private digital assets. This virtual machine allows for:
   a. *Persistent and secure storage*: based on a subset of the J2EE data access objects (DAO) pattern (see e.g. [7]) a persistency layer has been designed. Extensions allow for security on the basis of a PKI. There exist additional requirements for security in the management of data here. This extends to the security levels of commercial information systems. Up to now we neglect this here and leave this to future research.

b. *Information Exchange*: The J2EE transfer objects (TO) pattern (see e.g. [7]) is a basis for transferring secure assets between IDIs. Additionally, a groupware oriented functionality is specified in the virtual machine for building communities in multiple layers.
c. *Metadata and Ontology*: LPDAs are described at least using a fixed set of metadata that describes each LPDA by its technical parameters. These are name, format, owner, creation data, and several more depending on the input information given with the asset.

   An additional ontology is available as global information source in the IDIs PDAM system, for more details see [8]. This global information has two parts: A technical part describes relevant information of the IDI, i.e. infrastructure information that must be used by software for automatic conversion of LPDAs into a new IDI during the migration process.

   A personal global part is used for connecting LPDAs with the owners information background, e.g. helps understanding the relevance of individual LDPAs in the owners mental environment. The owner might have means to integrate ontological information to her or his individual LPDAs. This is provided by the specified virtual machine and can be used in the IDI.

2. *Proactive Agents*: A proactive agent is a software system continuously monitoring its environment and reacting to events according to predefined rules, see for instance [9]. We require a proactive agent monitoring each IDI managing LPDAs. Whether there should be one for all in a defined cyberworld, e.g. available as a web based service, or one for each IDI is up to architectural decisions.

   The proactive agent's responsibility is to monitor changes relevant to the LPDAs especially due to migration necessities and outdated signatures. Furthermore, the agent might have transactional groupware functionality for secure interactions between IDIs in order to fulfill some defined tasks ore goals, respectively. Therefore, the proactive agents look for events that are described in an event specification language.

   We use here an extended subset of ADL – a flexible and extensible activity description language for active information systems (cf. [10]). This is a rule-based language that allows for the specification of arbitrary events, for example changes in the environment of an IDI, detecting new IDIs or monitoring temporal aspects like outdating of information.

   In order to achieve an accurate event monitoring that suffices or goals for managing LPDAs each IDI or relevant system in the cyberworld must be active in the sense that it transmits necessary information into its environment, i.e. the cyberworld. Thus, in order to achieve our goals there are some requirements to be met by all systems involved in the cyberworld.

   Each proactive agent provides appropriate reactions due to detected events. The main actions taken in our long term preservation issues are asset migration and asset re-signaturing. The reactions are specified via the aforementioned activity description language, too, since it allows for flexible integration of services. The services detailed above are used as functions in the procedural bodies of ADL rules.

Figure 3 presents a bird-eyes view on the architectural inner structure of a component for preserving PDAM services within each IDI.

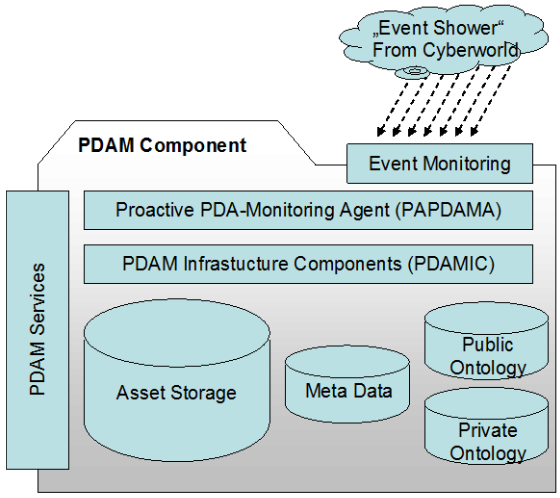


Figure 3: Architectural Aspects of a PDAM

## 3.3. Private Digital Asset Migration

The most vital activity of the proactive private digital asset monitoring agent (PAPDAMA) is to manage the migration of the assets into a new (future) IDI as soon as the prevailing infrastructure will be substituted. We want the complete information to be migrated into the new environment, that is the assets, the contingently updated meta data, and all ontological sources.

Prior to data migration the environment for managing all data has to be established into the new IDI. There exist two approaches: the first is that a PDAM-component is available for the new IDI by a vendor and can therefore be bought if not already provided by the IDI itself.

In the second approach not only the data are automatically transferred into the new IDI but also its managing environment, that is all services, components and agents. To accomplish this, the virtual machine approaches of [1] must be elaborated to a global extend: All possible future individual digital infrastructures must provide the functionality of such a virtual machine.

This can only be achieved by standardisation bodies: we postulate such standards in the future not only for the sake of the individual user but also for the sake of the human digital heritage in a global sense, see for example the ongoing activities of the UN supported CODATA organisation.

In both cases the PAPDAMA recognizes the new IDI and knows about the necessity to migrate all assets. This event can only be monitored if it is made explicit by the owner of the assets. Thus she or he must make the intention to move to the new IDI known to the PDAM explicitly.

At the time of this event both IDIs, the old one and the new one, must exist in the cyberworld in parallel, i.e. both infrastructures are up and running in parallel. On recognizing this event PDAM services based on the information exchange (TO) pattern will be executed. The latter transfer all assets and ontology information into the new IDI.

To accomplish this, each new IDI has to have services to transform the data structures of each asset, if this is necessary for the handling of the asset in its new environment. If the data structure of a signed asset is to be changed it must be re-signatured in this process. The meta data is transferred too, and additionally updated due to new technical information in the new IDI, e.g. new data format etc.

### 3.4. Re-Signaturing Digital Assets

Obviously, there exist points in time when the resignaturing of secure assets is necessary. From the descriptions above these points in time are at least the outdating of a signature and data structure manipulation during a migration process. Thus each asset is monitored in the PDAM with regard to these events.

Each asset affected by such an event must be resignatured *before* the existing signature of the asset is outdated, since thereafter there is no way back to a trustworthy status of this asset. Due to the reliable PKI procedures each new signature of an asset has to incorporate the old signature, too.

This results in a signature-procedures-stack that is part of the PDAMIC and has to be migrated from an old to a new IDI as well, if necessary. Figure 4 (cf. [11]) depicts the result of a one-time re-signaturing of a signed document using a PKI standard. More details on the secure management of LPDAs can be found in [11]. Verification procedures complete each PKI-based trustworthy asset management and result in additional data for each asset.

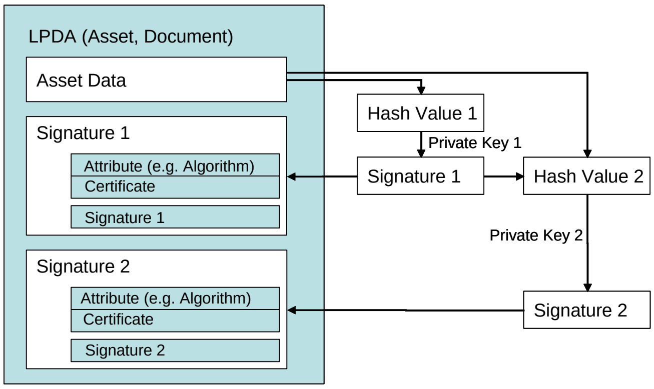


Figure 3: Result of Re-Signaturing

## 4. Prototypical Implementation

We established a prototypical private digital asset management (PDAM) system based on state-of-the-art web services technology that allows for secure asset storage. Furthermore, functionality has been integrated for asset interchange in order to support processes of the type presented in figure 1.

This prototype shows the feasibility of the PDAM architecture presented above on the basis of web services that implement the virtual machine. This prototype includes most of the security features including the usage of a public key infrastructure.

The prototype additionally provides a set of groupware functionality that supports group and community building. Secure asset interchange within the cyberworld is supported, too. The cyberworld is the whole internet up to now, i.e. all IDIs have to be part of the global net.

Asset management can be handled by each IDI on its own, i.e. each IDI provides all services on a local server, or can be delegated to other (central or distributed) PDAM servers (that are IDIs as well, in our notation). Due to strong authentication procedures the owner of an asset can delegate her or his asset management to another PDAM instance (IDI) while holding full control w.r.t. to asset handling.

A case study implemented automatic "re-signaturing" documents when a signature is going to be outdated or when the signaturing instance will become obsolete. Of course, this can only be accomplished when enough information about these events is provided by other IDIs or other members of the cyberworld.

Ontologies provide defined public and private terminology that has been attached to assets in our case. Up to now we used ontologies only for integrating user-friendly functionality, e. g. a query answering system regarding the semantic content of the asset pool. This helps the owner in the long run to remember her or his assets, e.g. those of the last century.

Whereas the functionality of a PDAM system could be implemented into a current IDI by our prototype – and within that a lot more functionality for user friendliness – the support for automatic migration of asset pools between different generations of IDIs using different operating systems and different asset specific software could not be tested up to now. It is difficult get an operating environment as a testbed where these different generations are running in parallel as IDIs.

## 5. Future Prospects

In order to go on with our research on migration processes we look forward to a simulation approach. A resilient model has to elaborated, that is the basis for tests of automatic ontology- and meta data-based migration of assets with conversions where necessary. This model will support us with more insight into essential functionality for future systems that are the basis of the digital information and communication infrastructure of individuals on of mankind.

In the long run we consider LPDAs as active objects, truly “living” in a cyberworld, perhaps with an own “will” to survive. The basis must always be an artificial habitat that is implemented as state-of-the-art information technology for individuals. As the technology evolves over time – here we have to think of centuries and millennia – the active objects must preserve their original content while adapting to new environments.

## 6. References

[1] R. A. Lorie: Long term preservation of digital information, in ACM/IEEE Joint Conference on Digital Libraries, JCDL 2001, Roanoke, Virginia, USA, June 24-28, 2001.

[2] nestor – Kompetenznetzwerk Langzeitarchivierung: Proceedings "Herausforderung: Digitale Langzeitarchivierung. Strategien und Praxis europäischer Kooperation", 2007.

[3] H. Jasper: Private Daten: Herausforderungen für das Datenmanagement, in Informatik bewegt: Informatik 2002, Schubert S. E., Reusch B., Jesse N. (eds.), LNI, Springer, 2002.

[4] R. Hüttl: Aspekte der Authentifizierung und Signaturen anhand der Erfahrungen eines produktiven Internetservice zur Archivierung digitaler persönlicher Dokumente“, in Informatik bewegt: Informatik 2002, Schubert S. E., Reusch B., Jesse N. (eds.), LNI, Springer, 2002.

[5] F. Jacobi: Virtuelle Maschine zur langfristigen Verwaltung von digitalen privaten Daten – Architektur und Infrastruktur, Bachelor Thesis, TU Bergakademie Freiberg, 2004.

[6] S. Goeser: Zur Langzeitpräservierung von digitalen Inhalten, Datenbankspektrum 3/2002, dpunkt Verlag, 2002.

[7] D. Alur, J. Crupi, D. Malks: Core J2EE Patterns: Best Practices and Design Strategies, Prentice Hall / Sun Microsystems Press 2001.

[8] B. Haberlach: Ontologien im Management digitaler privater Assets, Bachelor Thesis, TU Bergakademie Freiberg, 2003.

[9] S. Franklin, A. Graesser: Is it an agent, or just a program? A taxonomy for autonomous agents. In Intelligent Agents, Jennings, M., Wolldridge M., Springer, 1997.

[10] H. Jasper, O. Zukunft, H. Behrends: Time Issues in Advanced Workflow Management Applications of Active Databases, in Active and Real-Time Database Systems (ARTDB-95), Berndtsson M., Hansson J. (eds.), Workshops in Computing, Springer, 1996.

[11] P. Cyganski: Langfristige Sicherheit zertifizierte Dokumente, Bachelor Thesis, TU Bergakademie Freiberg, 2004.