A service for the preservation of evidence and data – a key for a trustworthy & sustainable electronic business

Steffen Schwalm[1]

Abstract: There is as high need to digitize and automatize business processes as well as to preserve the created electronic records for 2 up to 100 years or more to make business transactions evident against third parties. This requirement is not only mandatory for public administrations but also private companies especially in high regulated industries such as aviation, LifeScience & Pharma or financial sector. According to decade-long retention time it`s a challenge to ensure the availability as well as the preserve the authenticity, integrity, negotiability or reliability of electronic records. A digital archive service based on SOA and current technical standards to preserve all electronic records and their evidences provides a sustainable solution to this challenge. The paper shows based on current standards and the long-term experiences of the author possible solutions and an example of an architecture framework a digital archive service.

Keywords: long-term preservation, evidence preservation, eIDAS, SOA, records management, trust services, authenticity, integrity

1 Introduction

The utilization of the information technology for electronic business processes is established in public administration and private companies. Business records (a record contains any business relevant dataset together with its metadata [ISO15489]) increasingly exist in different digital forms and systems. At the same time, different national and international laws and regulations for the compliance of business processes and electronic records must be achieved e.g. [FDA], [FAA], [EASA], [eIDAS] etc. In summary, these regulations define in combination with corresponding standards that electronic records must provide their authenticity, integrity, negotiability, reliability and traceability to act as an evidence of the business transaction in which they were developed [KSH14], [To07] [ISO30301], [ISO15489]. To achieve these requirements, it`s not sufficient to preserve the verifiability of the authenticity and integrity of the records but also to be able to visualize them in 10, 20 or more years as needed. This means that a comprehensive digital preservation includes preservation of information and evidence of electronic records [KSH13], [KSH14]. A preservation services only focused on evidence preservation as currently discussed according to [eIDAS] is not enough if the user should be able to visualize, calculate or process his records in the needed way in after decades of retention time too.

[1] BearingPoint GmbH, Team Secure Information Management, Kurfuerstendamm 207-208, D- 10719, Berlin, Germany, steffen.schwalm@bearingpoint.com

These requirements are relevant for any business relevant data independently from the it-system in which they are created. This implies that a compliant digital preservation is a cross-sectional task and relevant for all records whose implementation as a SOA-service offers the utilization of synergy effects and avoids double efforts for the building up of archiving functionalities system by system. The paper presents an architecture framework for such an overall SOA-service for the digital preservation of any records. This example solution will be described based on the long-term experiences in different high regulated industries of the author as well as the state of the art legal, functional and technical requirements on digital archiving to preserve the information and its evidence value for any retention time.

2 Fundamental requirements on digital preservation of electronic records

2.1 Legal and organisational requirements

Like argued in the introduction: Due to legal requirements on electronic business, records management and the documentation of business transactions and decisions it`s necessary to preserve the records and linked transaction data for retention times up to 100 years whose start sometimes depends on an event in decades. Typical examples are licenses in the pharma-, aviation- or transportation industry which expires when the product will get out of service (e.g. Aspirin is under license since 1924) and the retention will not start before this event happens. In summary, the different legal condition requires the preservation of the authenticity, integrity, availability, negotiability and traceability of the electronic records to make this evident against third parties [KSH13] [KSH14]. The new [eIDAS]-regulation defines mandatory legal requirements on trustworthy electronic processes and documents by requirements mainly on

- Secure electronic identification
- Qualified trusted service providers concerning e.g.:
  - e-signature, e-seal and time stamps - the evidence for authenticity and integrity of electronic records
  - digital preservation – the basement for long-term verification of the named requirements on electronic records

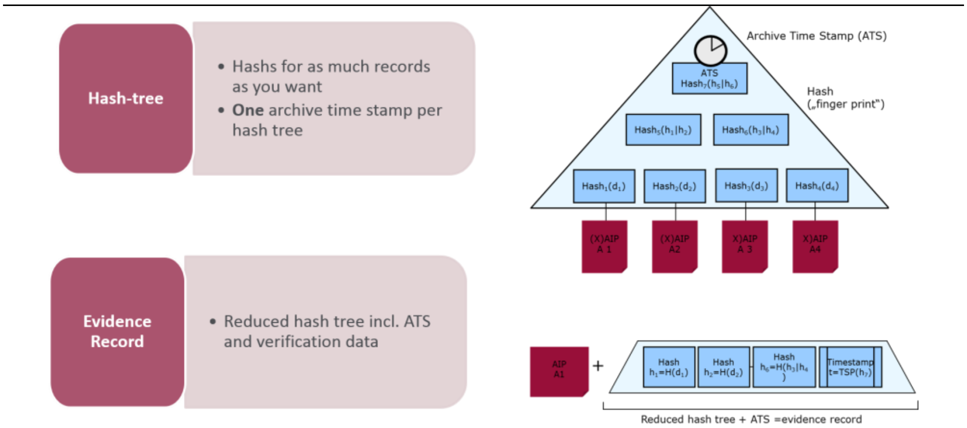
The implementing acts of [eIDAS] directly link to mandatory technical ETSI-standards. That means that technical standards become quasi legally mandatory for electronic records management. As one fundamental precondition, a professional records management ensures

- Well-defined and implemented roles, responsibilities and policies concerning reception, appraisal, systematic structuring and description, storage, preservation and distribution of records
- Reliability of the business processes based on organizational, technical measures
- Records controls, monitoring measures as well as well-defined processes for creation, capture and management of electronic records
- Information security requirements e.g. secure authentication, well-defined access rules, secure communication or data encryption
- Compliant signing of electronic documents based on [eIDAS] and current ETSI-standards
- efficient and standardized it-architecture independently from a special platform, modularized and so flexible concerning it-changes
- preservation of information and their evidences over the legally defined retention times

In a nutshell, a records management compliant to current standards e.g. [ISO15489] creates records which provides the evidence for business transactions against third parties, ensures so the trustworthiness of the records if they are needed and as a result a trustworthy information management [Wi15], [To07], [ISO30301].

The utilization of cryptographical measures e.g. qualified e-signatures, seals and time stamps enables public administration or private companies to preserve the evidence of their electronic records without losing the negotiability of the records. In [eIDAS] it`s legally defined that the trustworthiness qualified signatures and seals must be preserved which means the preservation of the evidence become manifested in the signature or seal. The evidence value of a qualified e-signature is the same as a handwritten signature, the seal makes the authenticity and integrity of the sealed record evident so the highest possible evidence value. These cryptographical measures are placed on the record that they become an inherent property of the records. That requires that measures concerning long-term preservation should focus on the record itself not the storage, the software environment etc. [KSH14], [KSH13], [KSHK15].

Currently the preservation of evidence is realized by resigning and rehashing of existing signatures and seals by signatures/seals of the same level and a qualified electronic time stamp before their security level expires [eIDAS], [KSH14]. The utilization of Merkle-hash [RFC4998], [RFC6283] trees ensures an efficient procedure by resigning a multitude of e-signatures or seals. The resigning contains all “old” signatures ad time stamps like illustrated in the picture below



Hash-tree

- Hashs for as much records as you want
- **One** archive time stamp per hash tree

Evidence Record

- Reduced hash tree incl. ATS and verification data

Fig. 1: Evidence preservation based on Merkle hashtrees

Concerning the fact that [eIDAS] is mandatory in EU and EFTA the utilization of e-signatures and seals will foreseeable increase. Regarding standardized and obligatory formats, requirements on the trusted service providers and their certification by the national authorities based on [eIDAS] the harmonization of a secure e-business and e-government is logical consequence. Especially the permissibility of mobile signatures, & server signatures together with the legal commitment to accept and verify any qualified e-signature, seal or time stamp from any qualified trusted service provider (a certified one) for all public administrations in EU+EFTA will provide new business options not only in high regulated industries and the increase of signatures, seals etc. as well as the need to preserve them to be compliant to European law. The current standards and standardization activities in evidence preservation related to [eIDAS] also consider the approach of evidence records e.g. [ASiC-E], [ETSI SR 019 510] as one possibility to preserve authenticity and integrity for long times.

The evidence preservation ensures the long-term verifiability of authenticity and integrity of electronic records. Furthermore it`s not only necessary to keep the reliability to the context of the business processes provable as well as the availability and negotiability of the records. This requires the preservation of electronic records in a self-contained way by addressing the needs for information- and evidence preservation. The evidence record based approach ensures an efficient and established implementation which is related to current standards on digital preservation e.g. [ISO14721] by connected standards [DIN31647], [DIN31644]. This enables a comprehensible digital preservation in combination of information- and evidence preservation.

2.2 Relevant standards

The picture below shows the main standards concerning preservation of information and evidence in a digital archive service compliant to the prior art.

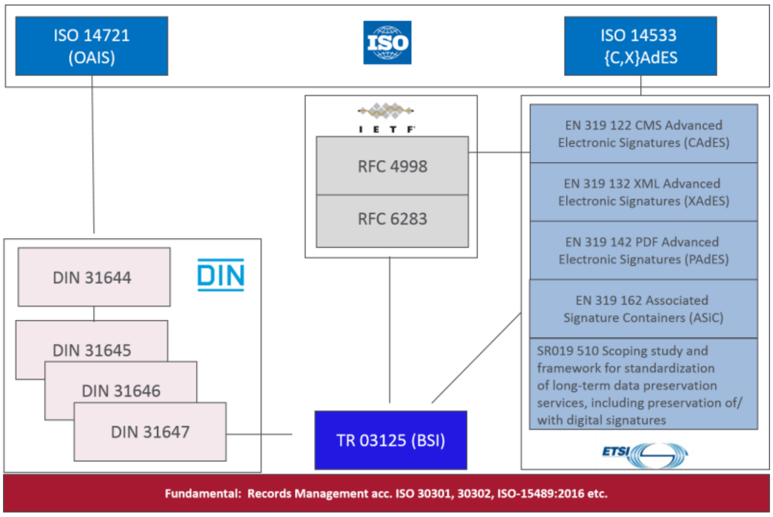

ISO 14721 (OAIS)
ISO
ISO 14533 {C,X}AdES
IETF
RFC 4998
RFC 6283
EN 319 122 CMS Advanced Electronic Signatures (CAdES)
EN 319 132 XML Advanced Electronic Signatures (XAdES)
EN 319 142 PDF Advanced Electronic Signatures (PAdES)
EN 319 162 Associated Signature Containers (ASiC)
SR019 510 Scoping study and framework for standardization of long-term data preservation services, including preservation of/ with digital signatures
DIN 31644
DIN
DIN 31645
DIN 31646
DIN 31647
TR 03125 (BSI)
ETSI
Fundamental: Records Management acc. ISO 30301, 30302, ISO-15489:2016 etc.

Fig. 2: Standards for holistic digital preservation

The main fundament for a trustworthy information management is a compliant records management based on well-defined and mandatory policies, responsibilities, tools and processes for capture, creation, structuring, storage and utilization of records in business transactions [ISO30301], [ISO15489], [ISO18829]. According to retention periods between 2 and 100 years whose start often depends on events in future the OAIS-model as well as the [ISO16363] or [DIN31644] defines the components and information packages for the preservation of the records themselves in a trustworthy long-term digital archive service. The standards are established worldwide. In combination with the [ISO14533] and the ETSI-Standards together with the [DIN31647] it`s possible to add the needed functionalities for evidence preservation to an OAIS-compliant digital archive service. Based on self-contained archival information packages acc. to [ISO13527], [PDF/A-3] or [ASiC-E] which contains the content, the metadata as well as the evidence relevant and technical evidence data such a long-term digital archive service ensures the availability, readability as well as the verifiability of the records until the end of the retention periods – independently from a special platform [KSH14], [KSHKW17], [Gia11]. Additionally, in EU and EFTA the ETSI/CEN-standards concerning e-signatures and evidence preservation should be noticed and use in a compliant digital archive service if the business transaction documented in records should be provable by third parties if they are needed, like defined in prior art e.g. [ISO15489], [ISO30301], [KSH14].

3 Architecture of a SOA-service for sustainable digital preservation

3.1 Functional cluster acc. OAIS

The main philosophy of SOA is to reuse existing it-services inside the current it-infrastructure as well as the ensure that new component built up for the archive are also reusable by other it-services and business it-solutions. Avoiding double effort to decrease the it costs is one main capability of SOA in state of the art it-infrastructures. Empirical the most functionalities used for digital preservation are not archive specific e.g. conversion which is used before a document is send out to external partners (word → pdf), signature creation and validation or integration platforms to connect different it-systems and services.

Like discussed in the chapters before a SOA-service for digital preservation should contain functionalities for the preservation of information and their evidences. It´s recommended to build up the components, modules and functionalities based on OAIS-model to avoid reinventing the wheel.

Processes	**Information packages**
Ingest	Submission Information Package
Data Management	Archival Information Package
Access	Dissemination Information Package
Archival Storage	
Preservation Planning	
System Administration	

Tab. 1: Processes and information packages acc. OAIS-model

A component in this context means the logical summary of associated modules and is here linked to one OAIS-process. A module contains the following properties:

- Contains distinct number of well-defined functions to fulfill distinct number of tasks
- Tiniest self-explanatory unit
- Logical as well as technical encapsulation possible (independency)
- Ability for orchestration of several modules to build complex processes
- Open and complete software documentation (supplier-independent service)

- Easy maintenance and module exchange

3.2 Ingest

The first component contains the modules for the SIP-transfer in to the digital archive service. Due to security reasons, it may be reasonable to encapsulate the functionalities for the connection of business it-systems to the SOA-services from the core ingest functionalities (e.g. conversion, validation etc.) in a so called Pre-Ingest-component.

To use the digital archive service as an overall SOA-service it`s necessary to summarize different standardized interfaces e.g. OASIS- and W3C-web service stack, SAP-ArchiveLink inside a connection module to integrate the records containing business-it. The ingest of electronic records possibly starts at the end of a business process when the records are closed or at the begin if extra sensitive records should be preserved early in the secure environment of a standardized long-term digital archive. This early archiving is typically used for document whose receipt should be verified and proved, signed documents or documents from a substitutional scanning especially in high-regulated industries e.g. aviation, banking, life sciences/pharma, government etc.

The connection module checks the access rights of the sending business it based on a secure authentication e.g. SAML, Kerberos and sends the submission or ingest request to the core-ingest-modules. Basically it´s recommended to limit the functionalities of the connection module to the following below [BSI TR-03125], [KSH14], [KSHK16]:

- Submission request (submission SIP and creation + storage of AIP)
- Extension of AIP (add new documents/data to existing AIP)
- Access data (whole packages as well as discrete access to single data)
- Access evidence record (reduced hash-tree acc. RFC 4998)
- Deletion of AIP (after retention time)

This means the connection module is used for Ingest and Access only differentiated by the different request – fully in the sense of SOA. With this approach, it`s possible to connect nearly every business-it to a digital archive service as a SOA-service also for SAP. For it-systems without standardized interfaces or where the effort for the connection to the archive is too high an upload-module inside the pre-ingest-component is recommended.

The core-ingest-component is based on defined and mandatory policies in which form, format, structure, with which metadata, evidence relevant data etc. the SIP from the connected business it or the upload-module are expected by the digital archive service. These technical rules (usually implemented based on a workflow engine) must be developed regarding the relevant legal and technical requirements for the digital preservation in responsibility of the data owner together with the it-department (or it-service provider). Based on these policies the SIP is checked and only after a successful validation the ingest process will continue. For this task format validation modules for metadata and content are used. but also, signature-/seal- or timestamp verification modules. Within the meaning of SOA, the crypto-module which is mainly needed for the evidence preservation component in the digital archive service is also used for signature-/seal-/timestamp verification during the core-ingest processes as well as the signature-/seal-/timestamp creation during the business processes before the data are send to the digital archive service [TR-3125], [DIN31647]. The named policies also define form, structure and content of the AIP created based on the SIP. Some of the needed more information preservation focused modules are:

- Conversion and validation of the original data e.g. to PDF/A (the original data can be stored in the AIP acc. to the business requirements)
- Definition of technical metadata acc. [PREMIS], validation of metadata regarding business context etc. (definition and addition Preservation Description Information)
- Mapping of IDs (ID for AIP to ObjectID from business it for access if necessary e.g. for SAP [KSHK15])
- Creation and validation of self-contained AIP-container which includes any metadata, content, evidence relevant and evidence data etc. necessary to achieve the business & legal requirements on digital archiving independently from a special it-platform (e.g. XML-based [ISO13527] or [ASiC-E], [PDF/A-3]) [KSHK16], [KSHKW17])

After the steps described before the AIP is correlated to the requirements for the preservation of information. So, it will be hand over to the evidence preservation component. To avoid reinventing the wheel it`s recommended to adopt the standardized reference architecture defined in [TR-03125] which is used cross-industry (e.g. aviation, government, health care, banking) and based on international standards.

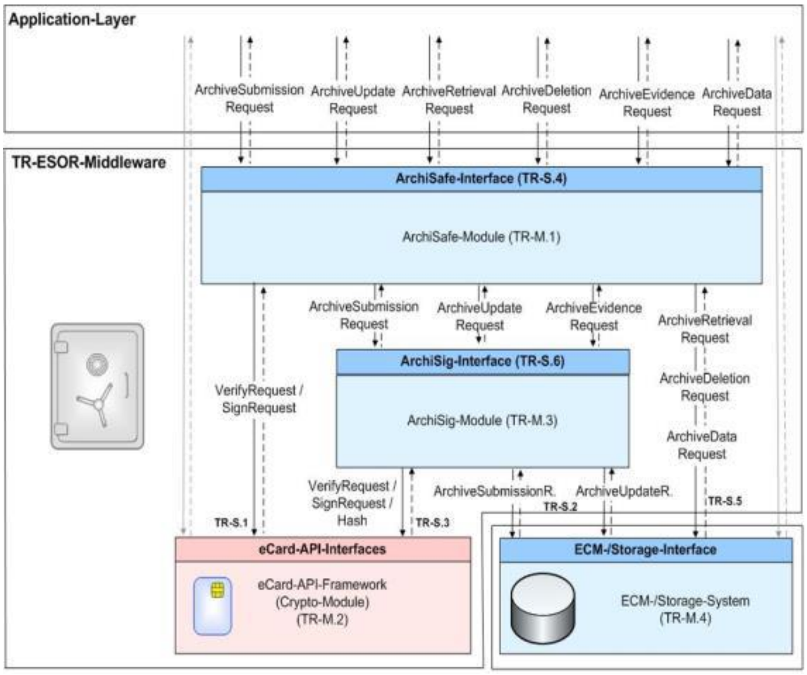


Fig. 3: reference architecture acc. [TR-03125]

The middleware comprises the core-functionalities for evidence preservation and is encapsulated from the rest of core-ingest by a secure ArchiSafe-interface acc. to a common-criteria protection profile [CC-PP-0049]. The AIP is given to ArchiSafe which checks the expected AIP-format and afterwards hand it over to the crypto-module for the obtaining of the evidence relevant data. The crypto-module will then send a response based on which the ArchiSafe-module will send the AIP to the ArchiSig- or RFC4998-module to add the AIP to the hash-tree which will directly be reduced to get the evidence records needed for external proves. This evidence record can be added to the not hashed parts the AIP. That means the AIP can be divided into evidence-record-relevant objects (ERO) or non-evidence-record-relevant-objects (NERO) like the definition of significant properties acc. to OAIS. Only the ERO will be included in the hash-tree-calculation. The evidence record will be stored in the NERO to create a completely self-contained AIP which the ArchiSig-module will create the AOID as a unique ID [KSH14], [DIN31647], [TR-03125]. The AOID will be handed over to the data management component as well as the business it to be able to access the archived records.

To avoid performance problems with complex AIP e.g. geographical data when XML-packages are in use for AIP it may make sense to store the content in a secure temporary storage and add the link to the AIP before it`s handed over to the evidence preservation component. During hash-tree-creation the ArchiSig-module can access the secure temporary storage with a secure communication channel e.g. VPN to create the hash.

Afterwards the content can be added to the ERO and NERO inside the AIP to get a fully self-contained AIP.

3.3 Data Management

In a digital archive service, the data management is usually a standard repository with functionalities for data and metadata maintenance including the AOID, full-text indexing, a workflow engine to link the other components e.g. ingest, preservation planning etc. Furthermore, common products also contain functionalities for retrieval, access, preservation and planning and sometimes the system administration although most of them are different logical components acc. to OAIS. All access to the evidence preservation and the archival storage component should be executed via the data management to reduce the complexity of the digital archive service.

3.4 Archival Storage

The AIP will be stored in the archival storage. In terms of a SOA and long-term digital preservation it´s recommended to use standard storage independently from a special platform or supplier together with a security concept concerning [ISO-27001] to avoid undesirable alteration of the AIP. This concept as well as the digital archive itself should be integrated in a holistic and sustainable information security management system regarding [ISO27001]. This approach avoids the utilization of expensive WORM-technology where its high dependence on a proprietary platform and a special provider. According the aim of digital preservation to archive records in self-contained AIP to be able to prove their authenticity, integrity, negotiability, reliability and traceability by third parties – so directly on the AIP itself and independent from the used storage – it`s a logical break to be standardized all the time and then store the AIP in a completely proprietary environment. Due to retention times between 2 and 100 years which often starts depended on an event in future the utilization of WORM it`s an incalculable risk for the record availability. Additional the preservation of information and evidence as well as the wide used extension of AIP imply a data explosion [Spi11], [KSH14].

3.5 Access

The Access cluster contains the modules and functions for retrieval, access and visualization of the archived records. In the most cases, these functionalities are mainly part of the leading business it connected to the digital archive. The digital archive is based on system access rights which means it only checks the access rights of the leading business it not the right of one single user – this is part of the access rights management of the business it itself. This approach avoids the rebuilt authorization schemes inside the archive. Only if the leading business it can`t be connected to the digital archive e.g. if it does not contain the needed interfaces the functionalities of the digital archive are used for retrieval, access and visualization e.g. for data from file shares. In practice, these modules are part of a repository which also manages the access rights for this well-defined use case.

3.6 Preservation Planning and System Administration

The preservation planning cluster contains the modules and functions for monitoring the technical improvement in algorithm, data-, container- and metadata formats, interfaces, architecture, software functionalities etc. which could influence the preservation of information and evidence of the preserved records. This includes the categorization, maintenance and control on specifications, a valid risk management as well as the assessment of the software tools in use. Based on these preparatory measures concrete actions e.g. re-signing or re-hashing, format migrations, changes of tokens for authentication or used interfaces can be carried out before they are out of service and the functional capability of the digital archive is put in question.

The System administration cluster contains the functional and technical administration of the digital archive which requires secure access to the relevant functionalities based on secure authentication (e.g. SAML, Kerberos, certificate token), limited to responsible users and oriented on the defined service model. There should be one platform to administrate the whole archive to facilitate the administration and decrease security risk. To protect communication channel and access a certified gateway e.g. by the Federal Office for Information Security (BSI) or comparable is often used for client server.

Both clusters are typically part of the repository in a concrete digital archive in practice.

3.7 Architecture of a SOA-based digital archive service

The compilation of the shortly described modules and functions before results mostly in the example architecture of a SOA-based digital archive service below. The picture gives and overview oriented on the structure of the OAIS-model. It does not contain details due to clarity reasons:

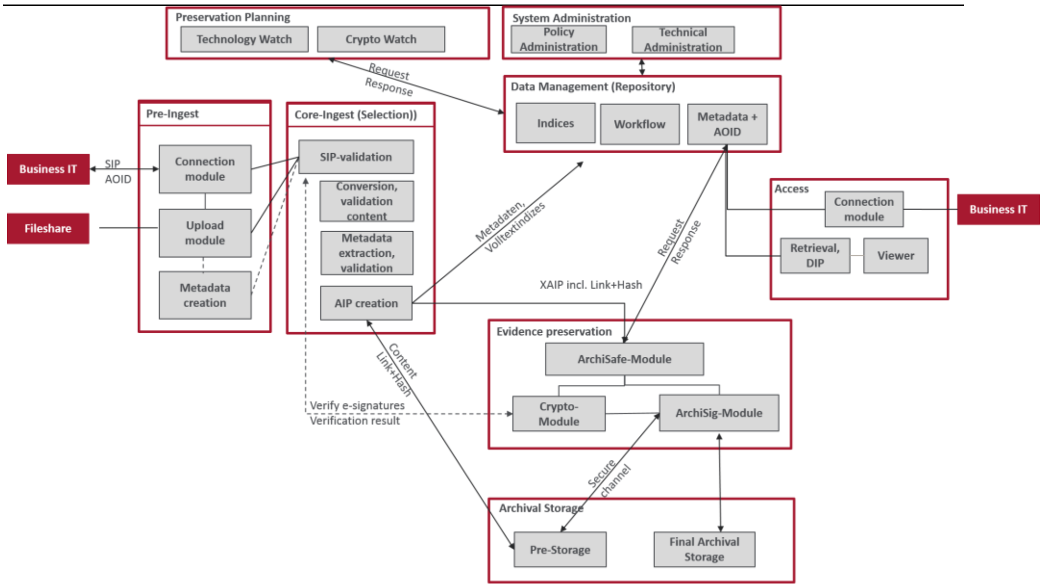


Fig. 4: Architecture Overview

4 Summary and Perspective

According to the new [eIDAS] as a EU and EFTA-wide consistently regulation the utilization of trustworthy electronic business process will foreseeable gain. Some examples for current solutions in this context are the implementation of complete digital application processes in the aviation, pharmaceutical or plant-protection industry. These processes include secure identification and authentication on a centralized European wide online portal serviced by the European regulation authorities e.g. EASA or EMA where the legally compliant signed applications are submitted as well as the whole approval process will take transparently combined with a secure communication between regulation authorities, applicants (industry) and stakeholders. Based on [eIDAS] and its trust services as well as eID-solutions the whole life cycle such complex business processes can be digitized in a secure and compliant way from the submission until the long-term preservation – but only if the preservation services will integrate both tasks – preservation of records and their evidences. A SOA-based digital archive service like describe above is so a key for a trustworthy and sustainable electronic business.

Bibliography

[ASiC-E] ETSI EN 319 162 – {1,2}, Electronic Signatures and Infrastrucutres (ESI); Associated Signature Containers (ASiC), ETSI (V1.1.1 (2016-024))

[CC-PP-0049] BSI-CC-PP-0049-2014. For Common Criteria Protection Profile for an ArchiSafe Compliant Middleware for Enabling the Long-Term preservation of Electronic Documents, Version 1.2 from Federal Office for Information Security

[DIN31644] DIN 31644:2012 Information and documentation - Criteria for trustworthy digital archives, 2012.

[DIN31647] DIN 31647:2015 Information and documentation - Preservation of evidence of cryptographically signed documents.

[EASA] EASA Part 21. Part-21 - Airworthiness and Environmental Certification

[eIDAS] REGULATION (EU) No 910/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC

[ETSISR019510] ETSI SR 019 510. Electronic Signatures and Infrastructures (ESI); Scoping study and framework for standardization of long-term data preservation services, including preservation of/with digital signatures. V0.0.4 (2016-12)

[FAA] Standard Airworthiness Certification Regulations Title 14, Code of Federal Regulations. PART 21—CERTIFICATION PROCEDURES FOR PRODUCTS AND ARTICLES

[FDA] CFR - Code of Federal Regulations Title 21. ITLE 21--FOOD AND DRUGS CHAPTER I--FOOD AND DRUG ADMINISTRATION DEPARTMENT OF HEALTH AND HUMAN SERVICES SUBCHAPTER A—GENERAL PART 11 ELECTRONIC RECORDS; ELECTRONIC SIGNATURE

[Gia11] David Giaretta: Advanced digital preservation. London 2011

[ISO13527] ISO 13527:2010, Space data and information transfer systems -- XML formatted data unit (XFDU) structure and construction rules, 2010

[ISO14721] ISO 14721:2012, Space data and information transfer systems — Open archival information system — Reference model, 2nd Edition, 2012

[ISO14533] ISO 14533: Processes, data elements and documents in commerce, industry and administration – Long-term signature profiles. 2014

[ISO15489] ISO 15489-1:2016 "Information and documentation - Records management – Part 1: Concepts and principles"

[ISO16363] ISO 16363:2012. Space data and information transfer systems - Audit and certification of trustworthy digital repositories. 2012

[ISO18829] ISO/DIS 18829:2015. Document management -- Assessing ECM/EDRM implementations – Trustworthiness. 2015

[ISO30301] ISO 30301:2011, Information and documentation - Management systems for records - Requirements. 2011

[KSH13] U. Korte, S. Schwalm, D. Hühnlein: Vertrauenswürdige und beweiswerterhaltende Langzeitspeicherung auf Basis von DIN 31647 und BSI TR-03125, Informatik 2013, GI-LNI, P220, ISBN 978-3-88579-614-5, S. 550-566, 2013

[KSH14] U. Korte, S. Schwalm, D. Hühnlein: Standards for the preservation of evidence and trust. Archiving 2014, S. 9-14. Springfield 2014

[KSHK15] U. Korte, S. Schwalm, D. Hühnlein, T. Kusber: Ersetzendes Scannen und Beweiswerterhaltung mit SAP. DACH-Security 2015. S. 72-85. Frechen 2015

[KSHK16] U. Korte, S. Schwalm, D. Hühnlein, T. Kusber: Beweiswerterhaltung im Kontext eIDAS - eine Case Study. DACH-Security 2016, Frechen 2016

[KSHKW17] U. Korte, S. Schwalm, D. Hühnlein, T. Kusber. B. Wild: Datenpakete zur Informations- und Beweiswerterhaltung – ein Vergleich. DACH-Security 2017

[PDF/A-3] ISO19005-3, Document management — Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 (PDF/A-3), 2012

[RFC3161] C. Adams, P. Cain, D. Pinkas, R. Zuccherato: Internet X.509 Public Key Infrastructure – Time-Stamp Protocol (TSP), IETF RFC 3161, http://www.ietf.org/rfc/rfc3161.txt, 2001.

[RFC4998] T. Gondrom, R. Brandner, U. Pordesch: Evidence Record Syntax (ERS), IETF RFC 4998, http://www.ietf.org/rfc/rfc4998.txt, August 2007.

[RFC6283] A. J. Blazic, S. Saljic, T. Gondrom: Extensible Markup Language Evidence Record Syntax (XMLERS), IETF RFC 6283, http://www.ietf.org/rfc/rfc6283.txt,

[Spi11] Stephan Spitz et.al.: Kryptographie und IT-Sicherheit. Grundlagen und Anwendungen. Wiesbaden 2011

[To07] Peter M. Toebak: Records Management. Ein Handbuch. Baden 2007

[TR-03125] Federal Office for Information Security (BSI): TR 03125 Version 1.2: Preservation of Evidence of Cryptographically Signed Documents (TR-ESOR).

[Wi15] B. Wildhaber: Leitfaden Information Governance. Zürich 2015