Long Term Preservation of Electronic records

Using Archive Interaction Protocol

Aleksej Jerman Blazic, SETCCE
Peter Sylvester, EdelWeb

Content

- Formal background
  - Archive definition
  - Formal requirements
- Technical background
  - Functional requirements
  - Infrastructure
  - Technology
  - Implementation

Motivation

- General
  - Preserving electronic heritage
- Specific
  - NASA; lost of Voyager mission data – OAIS as result
  - E-invoicing; proof of record existence for tax related processes – LTANS as result of previous work, namely DVCS, TAP, ArchiSig…)
  - Patenting; proof of intellectual property – LTANS as result
  - And many, many others…

Definition

- Archiving is a procedure of submission, retrieval, preservation, maintenance, professional management and usage of documented and archival material, which is not used for current usage until the expiration date according to formal and legislative requirements.

- Archive is a collection of records and documents that have historical, cultural or scientific value and are stored on physical media.

- Business related records treated as documentary material.

Formal requirements

- Electronic records are kept in their original form when:
  - Record data or record content is accessible and usable at any time,
  - Record data or content is preserved in the original form or in any other form that undeniably represents the original data sent or received,
  - The origin, time, location, sender and recipient of a electronic record or message is undeniably identifiable

Formal requirements

- Electronic records are kept in their original form when:
  - Technology and procedures used prevents any sort of modification, alteration or deletion of record data or content – integrity guarantee exists at any time,
  - Complementary data and means for security attributes (e.g. metadata, digital signatures) are preserved for the same archiving period as records
  - Procedures and means for extending the validity of security attributes are accordingly implemented.

Functional requirements

- Trusted archive service must
  - Provide evidence that can be used to demonstrate the integrity of data for the complete archive period
  - Demonstrate the validity of data for the complete archive period
- Trusted archive service accepts
  - Raw data
  - Signed data
  - Time stamped data (whatever that means)
  - Encrypted data

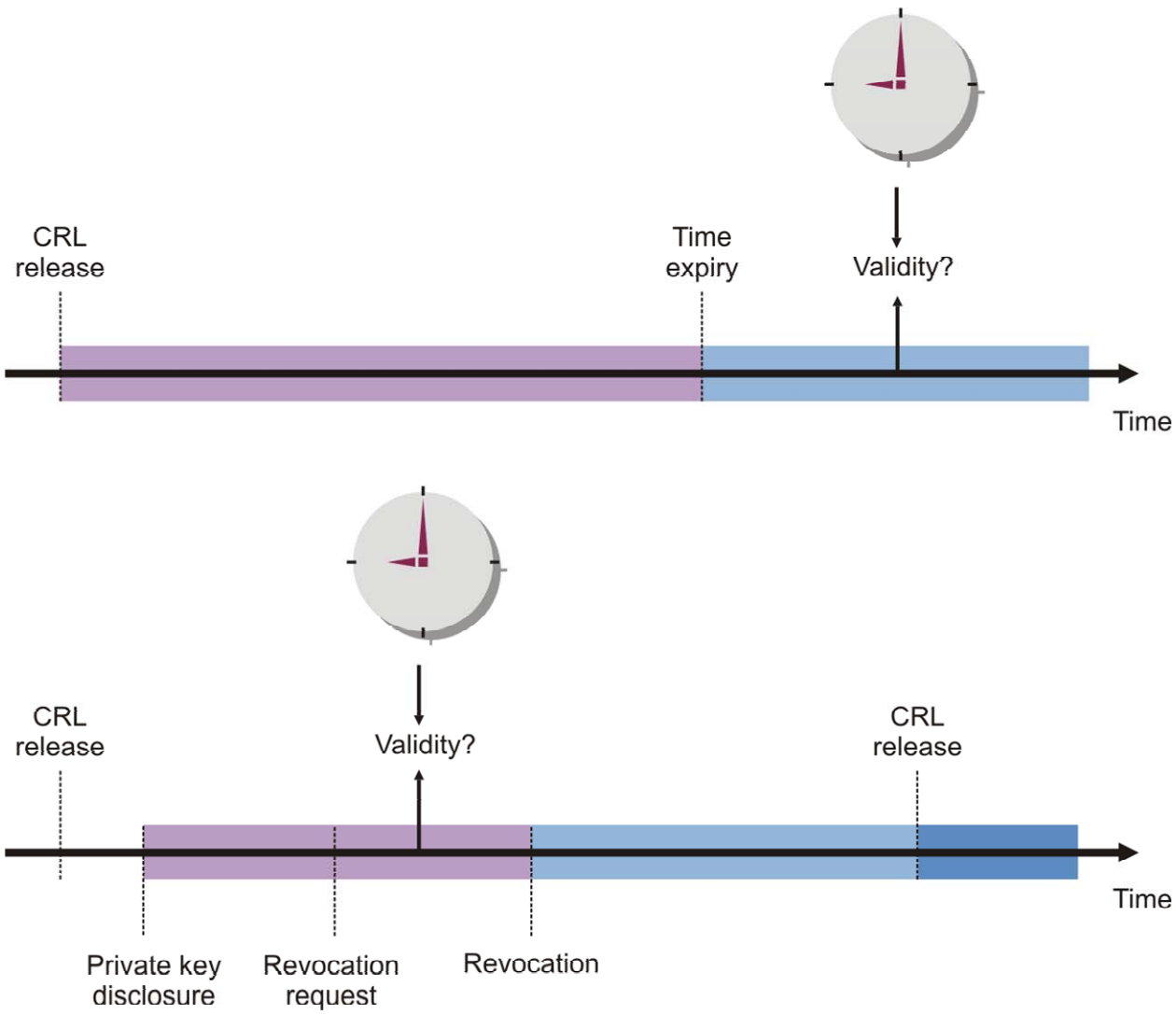
Functional requirements

- Trusted archive service must permit clients to request the following basic operations
  - Submit data objects to archive
  - Retrieve archived data objects
  - Delete archived data objects
  - Specify an archive period for submitted data objects
  - Extend or shorten the archive period for an archived data object
  - Specify metadata associated with an archived data object
  - Specify an archive policy under which the submitted data should be handled

Functional requirements

- Other requirements
  - Operate per a trusted archive service policy
  - Support data confidentiality
  - Transfer data and evidence from one service to another
  - Enable grouping and de-grouping of data objects
  - Support for large amounts of archived data objects processing
  - Support for long term validity of security attributes (digital signatures)

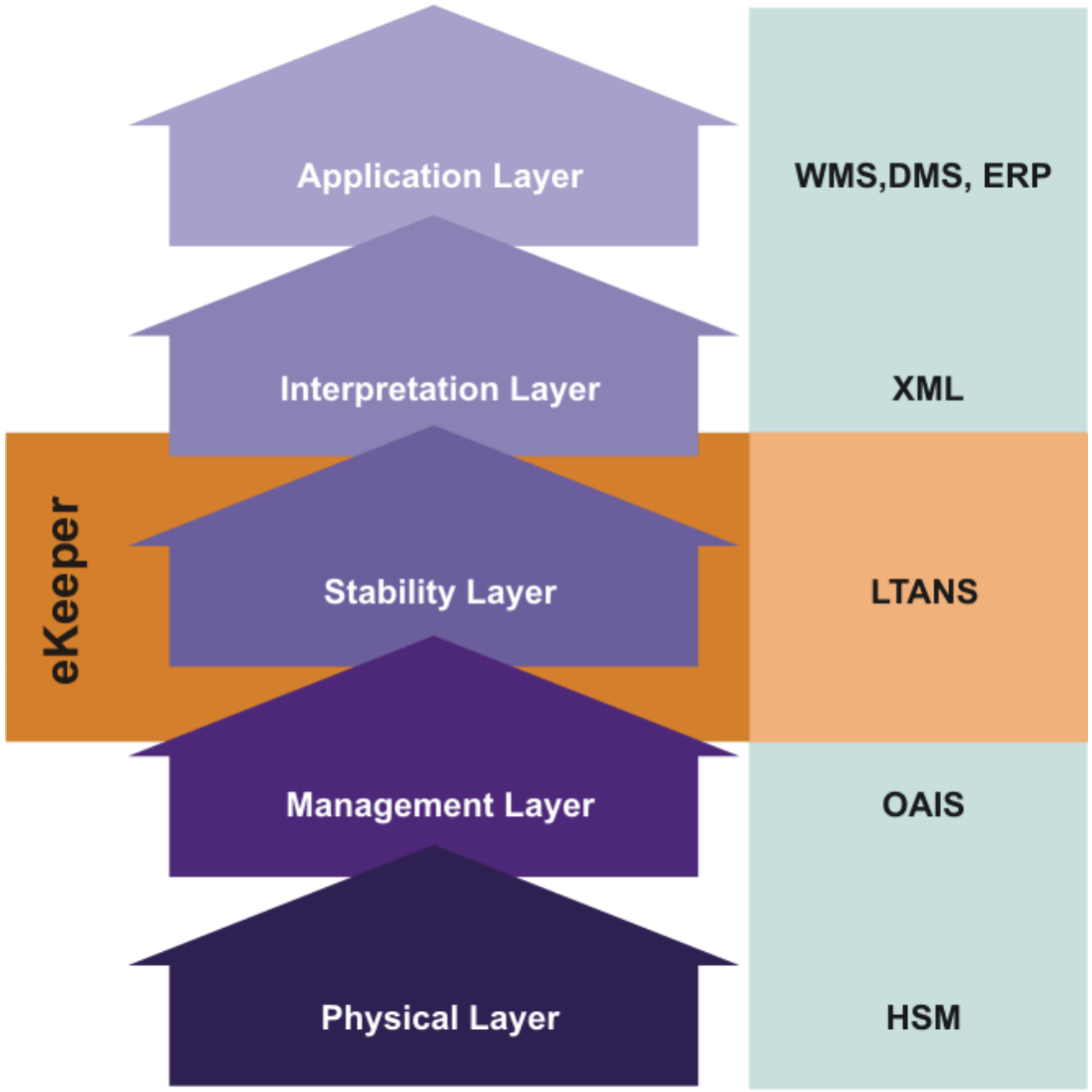
Long term validity



Trusted Archive Service

www.setcce.org

Archive service infrastructure



Trusted Archive Service

www.setcce.org

Physical infrastructure

- Basic infrastructure of trusted archive service
  - Interaction protocol
  - Archive objects
    - Data
    - Metadata
    - Digital signatures
    - Conservation attributes
      - Archive meta data
      - Complementary data
      - Evidence data
- Supporting infrastructure of trusted archive service
  - Communication network
  - Security mechanisms
  - Time stamping
  - Data storage or document management system

Interaction

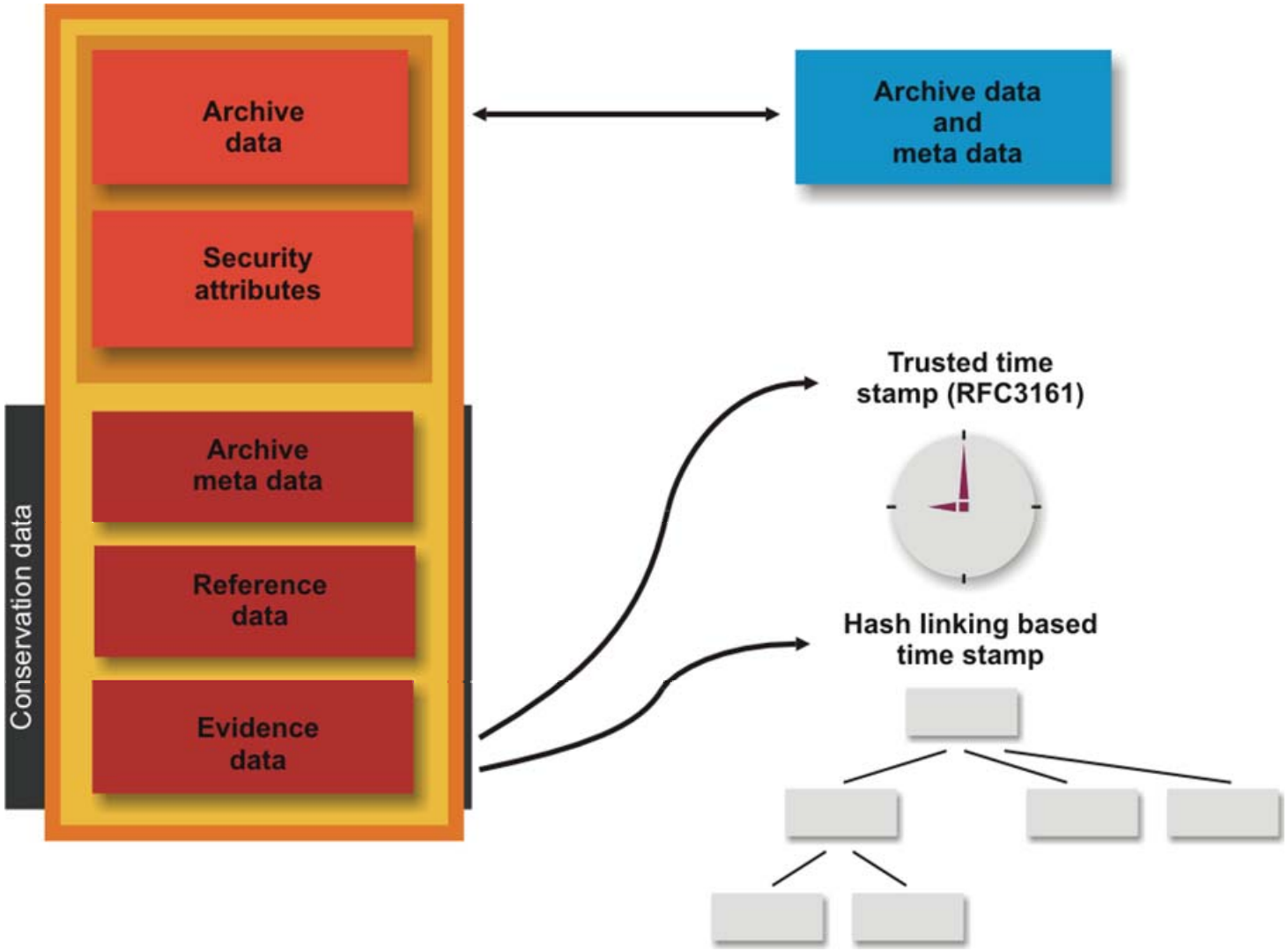
- Message based technical and formal interpretation of archive services
- Transaction based (asynchronous operation)
  - Request
  - Acknowledge (technical)
  - Response (business)
- Support for services

  archive/status/verify/export/delete defined by service
- Underlying authorization and transport services
  - SAML, SOAP, SSL, etc.

Archive object

- Logical building entities within trusted archive service
  - Archive data
    - Raw data
    - Metadata
    - Security attributes (digital signatures)
  - Conservation attributes
    - Archive meta data
    - Complementary data
    - Evidence data
- Long term maintenance of conservation attributes
  - Based on re-generation of evidence data
  - Introduces stronger security algorithms over requested archive period
- Logical structure
  - Physical presence of archive data
  - Archive data as reference or interpretation

Archive object



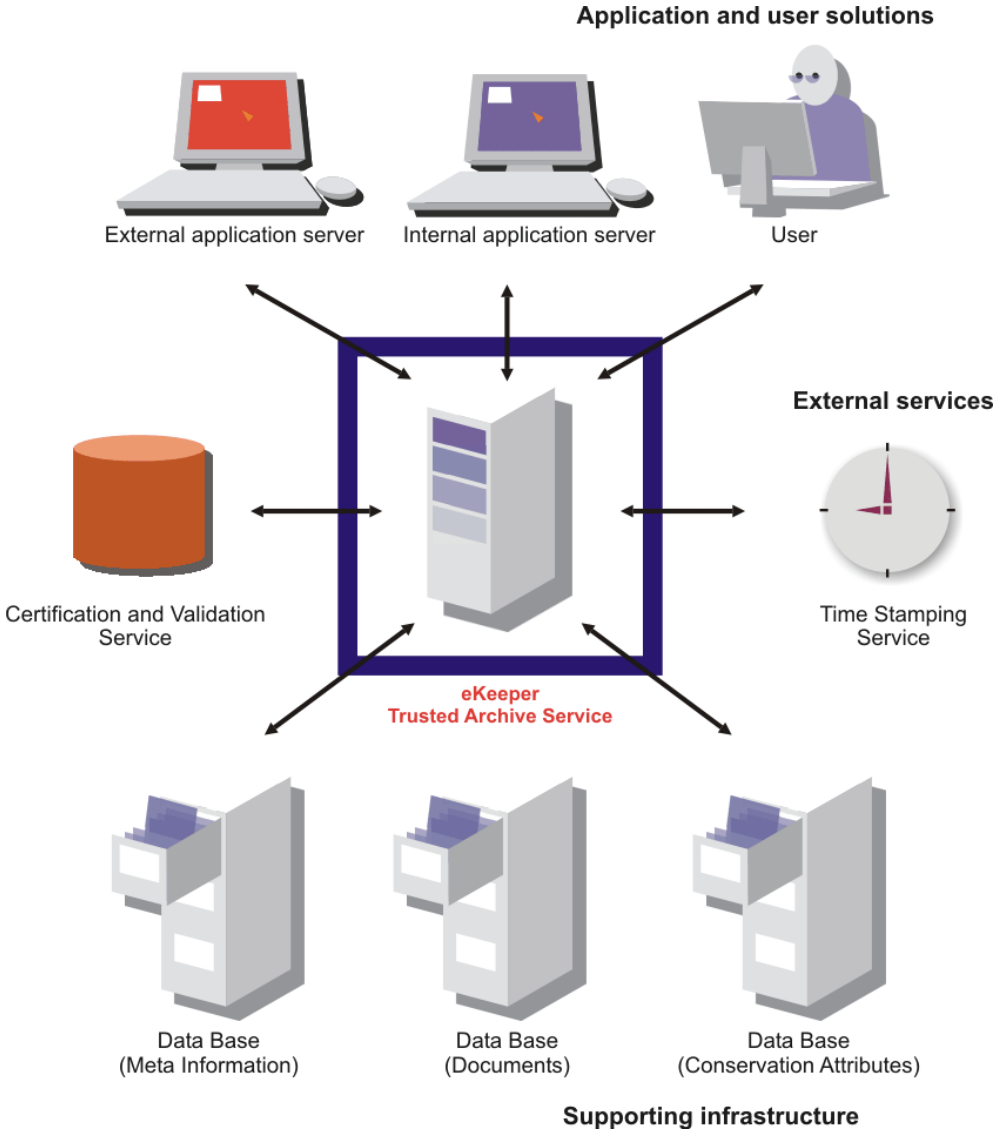
Evidence

- Trustworthy information (attestation) on
  - Archive data existence
  - Archive data integrity
- Based on trusted time
  - Time stamping
  - Binary or linear hash linking
- Grouping and de-grouping
  - Grouping – tree based hash linking with root value time stamping (e.g. Merkle tree)
  - De-grouping – redundancy values (leaves and nodes) used for calculating root value
- Preserving long term validity by reapplying evidence data over
  - Existing evidence data or
  - Existing evidence data and archive data (when security algorithms became insufficient) – requires archive data presence

Implementation

- Trusted archive service implementation
  - Basic service
    - Document storage not supported
    - Conservation attributes generation, storage and refreshing
    - Confidentiality
  - Advanced service
    - Document storage
    - Conservation attributes generation, storage and refreshing
    - Confidentiality optionally based on encryption mechanisms

Infrastructure



Trusted Archive Service

www.setcce.org

TAS in practice

- Implementation of trusted archive service
  - Second generation of trusted archive service
  - Accepts
    - Raw data
    - Signed data
  - Performs data grouping and de-grouping
  - Evidence records based on RFC3161
  - Redundancy operation supported
  - EDMS integration for business processes dematerialization
  - Demonstration service available on-line

    http://demo.setcce.org/ekeeper

Standards

- Implementation standards
  - Data structures
    - DoublinCore, IETF Atompub, ebXML registry…
  - Interaction
    - LTANS LTAP, W3C SOAP, ebXML …
  - Integrity
    - SHA1, SHA256, SHA384, SHA512, RIPEMD160…
  - Signatures
    - RSA PKCS#7, W3C XMLDSig, ETSI/W3C XAdES…
  - Evidence record
    - IETF RFC3161
    - Entrust XMLTS
    - LTANS ESR

Questions

Aleksej Jerman Blazic, SETCCE
aljosa@setcce.org

Peter Sylvester, EdelWeb
peter.sylvester@edelweb.fr