Klaus Stranacher, Bernd Zwattendorfer, Vesna Krnjic - "Secure and Efficient Processing of Electronic Documents in the Cloud" - Proceedings of IAIDIS International Conference e-Society, 2013, pp. 217-224

SECURE AND EFFICIENT PROCESSING OF ELECTRONIC DOCUMENTS IN THE CLOUD

Klaus Stranacher, Bernd Zwattendorfer, Vesna Krnjic
Graz University of Technology, E-Government Innovation Center, EGIZ
Inffeldgasse 16a, 8010 Graz, Austria

ABSTRACT

Electronic documents are often exchanged in e-Government and e-Business processes. In e-Government, the usage and importance of electronic documents has significantly increased particularly in cross-border scenarios, especially due to the implementation of the EU Services Directive. To ensure genuineness, in many situations electronic signatures are applied on the documents exchanged. Besides of their application, verification of electronic signatures is essential. Current solutions for the verification of electronic signatures usually support a subset of existing signature formats only. In addition, electronic documents require some kind of detailed description on higher level, e.g. through meta data. If corresponding meta data are recognized as incomplete or wrong during document exchange, additional costs and time delays may occur. Here, the need of previous data validation arises. To overcome these issues, we introduce an approach for secure and efficient processing of electronic documents, particularly focusing on signature and meta data verification. Our solution follows a generic concept and is not limited to certain use cases. Nevertheless, we present our approach based on the findings of the EU Large Scale Pilot Project SPOCS. Finally, we elaborate on the movement of verification and validation services into the cloud.

KEYWORDS

Electronic Documents, Signature Verification, Data Validation, Cloud Computing

1. INTRODUCTION

Electronic documents are important parts of most e-Government and e-Business processes. Their significance particularly increased with the progressing **implementation of the EU Services Directive** (European Union, 2006). The main objective of the directive is to establish a framework for easily setting up and exercise a service in another EU Member State by using **electronic procedures. Here, electronic documents are one of the key enablers to achieve this goal**. To guarantee authenticity and integrity of electronic documents, usually electronic signatures are applied to electronic documents. The validity of an electronic signature can be unambiguously determined by the receiver of a signed document using signature verification services. Current existing signature verification services are limited to verify only certain signature formats. In general, they support several kinds of standard signature formats as defined by the European Commission Decision on *"establishing minimum requirements for the cross-border processing of documents signed electronically by competent authorities under Services Directive"* (European Commission, 2011). In addition, they are able to verify a few other formats suitable for their field of application. Nevertheless, there exists a lack on comprehensive signature verification services supporting a wider field of signature formats.

In addition to the verification of genuineness of electronic documents, automatic processing of electronic documents is essential for a cost reducing, time saving, and efficient public administration. The basis for automatic processing is availability of machine-readable data, i.e. structured electronic documents and appropriate meta data. Nevertheless, additional costs and time delays may arise if electronic documents or meta data are recognized as incomplete or wrong. Here, the need of a previous data validation arises.

The European Large Scale Pilot SPOCS[1] developed an electronic document interoperability framework and document container format, called OCD - Omnifarious Container for eDocuments (SPOCS, 2011). Among other things, this framework defines how to verify and validate an OCD container and all affiliated electronic documents. The focus of the OCD interoperability framework has been given on the verification of the electronic signatures applied to the container and the contained electronic documents. In this paper we present and propose mechanisms for secure and efficient processing of electronic documents.

The remainder of this paper is organized as follows. In Section 2, we describe the electronic document interoperability framework and the Omnifarious Container for eDocuments (OCD). Additionally, we point out existing solutions for verifying and validating electronic documents. Section 3 elaborates on external signature verification services to support an extended set of signature formats. In addition, a data validation mechanism is proposed. These verification and validation facilities base on the OCD but are not limited to this use case. The subsequent Section 4 elaborates possibilities to transfer verification and validation services into the cloud. Finally, we draw conclusions including an evaluation of our proposed solution and discuss future work.

2. RELATED WORK

Basically, electronic documents can be divided into structured, unstructured, and container formats. The content of structured document formats follows a well-defined schema and is therefore machine-readable and can be easily processed. The most popular structured eDocument format is XML. In contrast, unstructured electronic documents, such as the PDF format, cannot be automatically processed. They are mainly used for visual representation of document content. Container formats specify how different types of data are stored in one container. Additionally, all required information, which third parties would need for processing the documents, is stored in the container. One of the first container formats was MIME[2]. In the meanwhile, formats such as Open Document Format[3] (ODF) and Office Open XML[4] (OOXML) have increased in popularity.

Looking at the e-Government landscape in Europe, every country has its own eDocument infrastructure deployed based on existing standards and technologies. Many national applications are using XML-based specifications for information and document exchange. However, national XML specifications cannot be automatically processed nor automatically interpreted by any third party without the knowledge of the schema of the particular document. Due to the EU Services Directive the need of interoperability for electronic documents, especially on a cross-border level, has significantly increased. This need for interoperability has also been discussed by Rössler and Tauber (2010).

The challenge on interoperability has been taken up by the European Large Scale Pilot SPOCS. Here, an interoperability concept has been introduced, which bases on the individual national infrastructures of the participating EU countries and builds an interoperability layer on top of it. This concept is called Omnifarious Container for eDocuments (OCD) and represents an interoperable multi-layer framework for cross-border exchange of electronic documents. The container supports all formats and technologies of electronic documents and is easily extendable to support new formats and technologies too. Additionally, semantic interoperability and authentication mechanisms for guaranteeing the authenticity of an OCD container are provided.

The specification of the OCD container (SPOCS, 2011) consists of a logical and a physical structure. Thereby, the logical structure consists of a payload layer, a meta data layer, and an authentication layer. The payload layer stores all kind of electronic documents, which should be transported in the OCD container. To

[1] SPOCS (Simple Procedures Online for Cross-border Services) is an EU co-funded project out of the EU ICT Policy Support Programme and aims to overcome the obstacles raised by the EU Services Directive. http://www.eu-spocs.eu/.

[2] MIME (Multipurpose Internet Mail Extensions) are extensions of the standard RFC 822 and defined in RFC 2045, RFC 2046, RFC 2047, RFC 2048 and RFC 2049.

[3] ODF is a standard developed by the standardization organization OASIS and is specified in ISO/IEC 26300.

[4] OOXML is a standard developed by Microsoft and is specified in ISO/IEC 29500.

support automatic processing, the meta data layer has been introduced on two levels. The first level describes each payload document, while the second level describes the container itself. In addition to the signed payload documents, the whole container can be signed as well. This authentication layer is optional and enables the support of authenticity of OCD containers.

Two different physical structures are defined to implement the logical structure. The ZIP based OCD relies on the ETSI specification on Associated Signature Containers - ASiC (ETSI, 2012) and uses XAdES signatures (ETSI, 2009) for the authentication layer. This ZIP based OCD is primarily suitable for back office applications. The second structure is a PDF based OCD where the master PDF represents the meta data and the payload documents are added as attachments. Here, PAdES signatures (ETSI, 2010) are used for the authentication layer. This technology is especially suitable for applications where citizens are directly involved.

To handle OCD in real live scenarios, operations on the core elements of OCDs are defined. The OCD Creation method defines how an OCD container is created. As input, this method takes arbitrarily signed or unsigned electronic documents with appropriate meta data. The resulting OCD container can be signed optionally. The OCD Validation and Verification method defines how an OCD container is validated and how signature verification is carried out. This method takes an OCD container as input. The output of this method represents the corresponding validation and verification report. The described methods have been implemented as open source software modules and are freely available for download on Joinup[5].

In addition to OCD, several other signature verification activities have been established. The European Commission published a tool, called SD-DSS[6], which is capable to verify signature formats based on the European Commission Decision on standard signature formats (European Commission, 2011). Furthermore the EU Large Scale Pilot PEPPOL[7] addressed issues concerning the signature verification in the field of e-Procurement and developed a suitable signature verification service[8]. Nevertheless, these services do not support verification of national and proprietary signature formats.

3. VERIFICATION AND VALIDATION SERVICES

Verification of the genuineness of electronic documents is important to trust the authenticity and data integrity of these documents. Usually, electronic signatures are the means of choice for guaranteeing authenticity and integrity. Verification of these signatures is essential for their further processing. In addition, data validation, i.e. the validation whether the present data are appropriate and correct or not, gains more and more importance. Both, signature verification and data validation are necessary for a secure and efficient processing in e-Government or e-Business scenarios.

The following sub-sections elaborate on signature verification and data validation of electronic documents incorporating external verification and validation services. Thereby, we concentrate on the use cases related to the EU Services Directive and the implementations of the large-scale pilot project SPOCS, focusing on OCD container verification and validation. Nevertheless, our approach is not limited to these use cases and applies for all processes where electronic documents are involved and must be processed. In addition, we show external dependencies to our methods, which are able to be outsourced to cloud computing, enabling high scalability and cost savings.

3.1 Signature Verification

For electronic signatures various data formats exist. On the one hand, there are signature formats which are tightly bound to specific document formats, such as PDF signatures. On the other hand, there exist signature formats which can be used with almost every document format, e.g. XML and XAdES signatures. Based on the EU Services Directive (European Union, 2006) the European Commission established minimum requirements for the cross-border processing of documents signed electronically by competent

[5] http://joinup.ec.europa.eu/site/spocs/eDocuments/
[6] http://joinup.ec.europa.eu/software/sd-dss/home/
[7] PEPPOL (Pan-European Public Procurement Online), http://www.project.peppol.eu/
[8] http://www.peppol.eu/peppol_components/esignature/esignature

authorities under the Services Directive. Article 1 (1) of this decision defines three signature formats, namely *"XML or a CMS or a PDF advanced electronic signature in the BES or EPES format"* (European Commission, 2011), as minimum or standard formats to be processed by EU Member States. In addition, Article 1 (2) states that *"Member States whose competent authorities sign the documents referred to in paragraph 1 using other formats of electronic signatures than those referred to in that same paragraph, shall notify to the Commission existing validation possibilities that allow other Member States to validate the received electronic signatures online, free of charge and in a way that is understandable for non-native speakers […]"* (European Commission, 2011).

Actual existing signature verification services are limited to verify certain signature formats. In general, they support the standard signature formats and a few other formats suitable for their field of application. So usually national and proprietary formats, as mentioned in Article 1 (2) of the EC Decision, are not supported. Stranacher and Kawecki (2012) presented a signature verification service, which introduced the concept of external signature verification services. This service can be used to integrate the verification of national and proprietary signature formats. Their concept bases upon the OCD Validation and Verification Module but lacks on a concrete implementation of this mechanism.

In Figure 1 we show the concrete mechanism to integrate external signature verification services on the basis of the OCD Validation and Verification Module. As an OCD container can be signed itself and usually contains signed electronic documents, the container signature and the document signatures are divided. A format detection unit analyzes the signature and recognizes the signature format. The verification of standard signature formats is covered by the internal signature verification unit. National and proprietary formats are verifiable via external verification services. These external services have to be defined in the configuration of the module. Within the configuration, a mapping between the MIME type representing the signature format and the respective external service is given. Based on this mapping, the verification of the national and proprietary formats are outsourced to the external service via a connector. This connector creates the request to the external services and receives the corresponding response. Additionally, the connector converts and transforms the response into the OCD module internal verification result format. Finally, the result generator unit collects all results, including additional validation results from a basic validation[9] (not shown in the figure for clarity), and generates an XML based verification report.

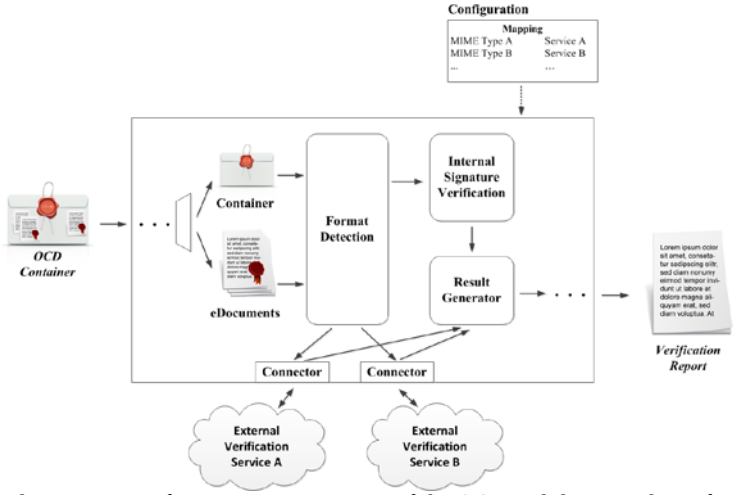

Figure 1. External signature verification services as part of the OCD Validation and Verification Module

Examples for external signature verification services supporting national and proprietary formats are the:

- Lithuanian verification service: This service supports the verification of the ADOC format specified by Director General of the Lithuanian Archives Department (2009).
- Austrian verification service: This service supports the verification of the PDF-AS format. PDF-AS is a proprietary Austrian format based on PDF and explained by Leitold et al (2009). Based on the solution of Zefferer et al. (2011), a Web-Service of this service based on SOAP will be available soon.

[9] The basic validation validates if the OCD container is compliant to the OCD specification.

These verification services may also be maintained within a cloud. An evaluation of this cloud-based approach is given in section 4.

3.2 Data Validation

Electronic documents are usually received by a service or application to be used for further processing. For instance, a public authority receives a request for opening a business and forwards it to the relevant competent authorities, which actually handle the request. If these documents are recognized as incomplete or wrong during further processing, the entire process must be stopped. To avoid associated costs and time delays, which may occur in such situations due to the necessity of manual interactions, a previous automated data validation is necessary. Data validation simply means that the data is verified if it fulfills the requirements for the subsequent process.

Basically, two different kinds of data can be distinguished. On the one hand, *meta data* provides information about the accompanied data such as the creation date or the creator of the data. Usually, meta data is available as machine-readable data. For instance, the OCD container comprises a meta data layer which describes the container itself (so-called meta data level 2) and the included electronic documents (so-called meta data level 1). On the other hand, the OCD specification (SPOCS, 2011) defines *document data* as a unified and machine-readable description of the content, optionally including the real content data. This document data introduces a mechanism to describe the content of an electronic document, which is available in a non-machine-readable format only[10], but still in a structured way though.

Document data defines a set of information on the level of electronic documents for storing machine-readable content. This set of information includes:

- A type identifier, which indicates the type of the document data, e.g. this is a birth certificate.
- A description of the structure, e.g. a birth certificate must contain the name and date of birth of the person as well as the names of her parents.
- The extracted values out of the original electronic documents satisfying the defined structure, e.g. the real name and date of birth of the person as well as the real names of her parents.

Figure 2 shows the basic principles of meta data (a.) and document data validation (b.). For meta data validation, meta data to be validated serve as input for the validation. For instance, such meta data can be extracted from an OCD container. In addition, a meta data profile ID selects a certain pre-configured meta data profile. Such a profile defines the meta data structure, i.e. which meta data must be present (e.g. meta data must contain a sender and a subject) and optionally which content must be present in the corresponding meta data fields (e.g. the sender must be “John Doe”). Based on this profile, the meta data is validated. First, the structure of the meta data is validated. In the second and optional step, the contents of the meta data fields are validated against the selected profile. Based on these validation steps, a common validation result is generated.

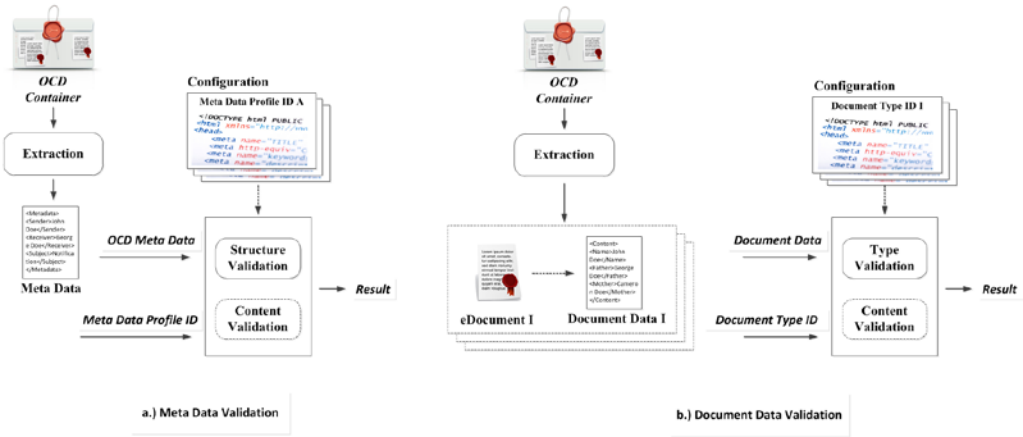

Figure 2. Meta data and document data validation

[10] E.g. a PDF document containing a scanned copy of a birth certificate

Document data validation is carried out according to a similar principle. Document data to be validated serve as input. For example, document data can be extracted from an OCD container. As second input, a document type ID is given, selecting a pre-configured document type profile. Thereby, a document type profile indicates the structure and optionally the contents of the document type (e.g. a birth certificate). Subsequently, document data are validated checking compliance against the profile, i.e. the data represent the given document type and – optionally – contain the required content. Finally, a validation result is generated.

Both meta data and document data validation base on XML schemata. During the validation process the data are verified if it is compliant to the given XML schema.

Figure 3 shows the integration of meta data and document data validation based on the OCD Validation and Verification Module. Here, meta data validation is an internal part of the module as the meta data scheme is OCD specific. Nevertheless, the concept of the proposed meta data validation is adaptable and can be used in various scenarios where validation of meta data is necessary. In addition, the validation of document data is linked to an external service as this validation is not OCD specific and thus follows a universal approach. Finally, the results of the meta data and document data validation are added to the verification report. External document data validation may also be maintained within a cloud. Section 4 elaborates on a possible cloud-based approach.

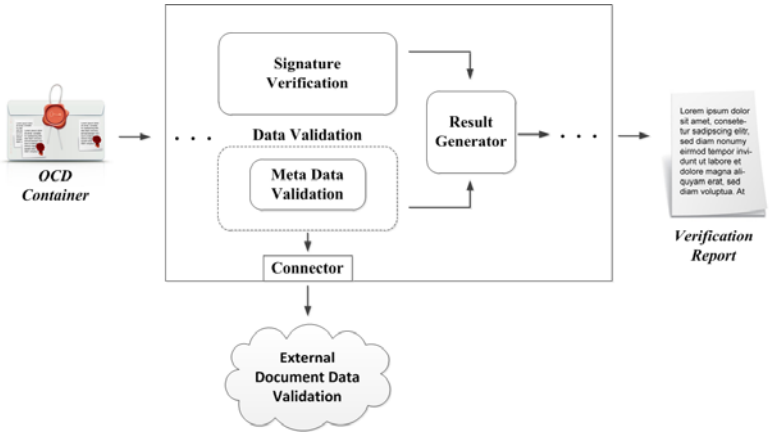

Figure 3. Data validation as part of the OCD Validation and Verification Module

4. VERIFICATION AND VALIDATION SERVICES IN THE CLOUD

Verification and validation of both electronic signatures and corresponding data applied to electronic documents are essential. In this section, we elaborate how cloud computing can be taken up to further improve verification and validation services of electronic documents.

Cloud computing (Mell and Grance, 2010) is currently on of the dominant topics within the ICT sector. The main aim of cloud computing – providing IT resources such as computing power or data storage based on a pay-as-you-go model – promises a lot of benefits. For instance, IT resources can be consumed just on demand and only effectively consumed resources are charged and must be paid. On the one hand, this provides high scalability for online services because required resources can be easily added. On the other hand, due to the flexible pricing model a lot of costs can be saved. By taking up cloud computing also for verification and validation services, these services can also take advantage of higher scalability and cost savings.

Basically, we see two main scenarios where cloud computing can particularly help improving verification and validation services, namely by deploying a

- Single external verification or validation service in the cloud or
- Brokered external verification or validation service in the cloud

We elaborate both approaches in more detail in the next sub-sections.

4.1 Single External Verification or Validation Service in the Cloud

Involving external verification services extend internal signature verification mechanisms, as the applied signature format might be proprietary and hence not supported by an internal service. This especially holds for country-specific signature formats, e.g. the Austrian and Lithuanian signature verification services as mentioned in Section 3.1. The verification of electronic signatures constitute a frequent process, hence such a national signature verification service may face a lot of requests. In particular, the amount of requests to be processed may not be constant. I.e. situations may occur where such national services have to cope with a high load. In such situations, simple verification services may not be able to handle load peaks and may tend to break down. More severely, in a worst case this can lead to a denial of service.

To bypass such bottlenecks, the verification service could be easily deployed as cloud service. The cloud guarantees nearly an independent amount of resources. Hence such potential bottlenecks could be easily overcome. In addition, applying the cloud computing paradigm offers some cost savings potential, as only the consumed amount of resources has to be paid. An imaginable scenario would be the implementation and deployment of a central cloud service per country, which is capable of processing individual signature verification requests. A similar approach, where countries host single and central gateways for individual data processing, can be found in the European Large Scale Pilot Projects STORK[11] and epSOS[12].

While scalability issues can easily be solved by applying cloud computing, the use of the cloud might bring up other issues in terms of security or privacy (Zissis and Lekkas, 2012). Before deploying such verification and validation services in the cloud, a thorough analysis on the cloud model to be applied is required. While public clouds offer the highest cost savings potential, private or community clouds might be favored as they allow higher control on the data to be processed (Catteddu and Hogben, 2009).

Finally, applying such a model is not limited to signature verification services. Needless to say that data validation services could follow such an approach too.

4.2 Brokered External Verification or Validation Service in the Cloud

While single external verification services in the cloud bypass the issue on scalability, they still leave the issue on heterogeneity of external verification services unresolved. Applying the single external verification services in the cloud model can lead to situations, where verification modules still have to manage several different interfaces to those external services. In other words, verification modules must support and implement the interface for connecting to the Austrian verification service, the interface to the Lithuanian verification service, etc. Such a model does not perfectly scale, hence we propose a brokered external verification service in the cloud similar to the brokered approach described by the Cloud Security Alliance (2011) as a second option. In this model, the verification module needs to support one interface to an external verification service only, namely to the brokered external verification service in the cloud. In addition, the brokered external service incorporates several other external verification services interfaces, e.g. the interfaces of several countries. In other words, such a service acts like a broker or hub between the verification and validation module and several external services. Summarizing, this approach provides two main advantages. The first advantage is scalability as the service is deployed in the cloud. The second advantage is the support of individual other external verification services to avoid heterogeneity.

However, this approach has also to deal with privacy and security concerns. Probably, private companies might take up this approach and hence data could be processed in a public cloud, which provide a lower security or privacy level. To bypass these concerns, it might also be feasible that the European Commission itself sets up such a service. Hence, to ensure higher control on the data to be processed such a scenario relates more to the application of a private or community cloud. Again, this approach is valid for both signature verification and data validation services.

5. CONCLUSIONS

[11] STORK (Security Across Borders Linked), https://www.eid-stork.eu/

[12] epSOS (Smart Open Services for European Patients), http://www.epsos.eu/

Secure and efficient processing of electronic documents and its affiliated meta data are crucial requirements for efficient and security-sensitive applications. Our described approach shows solutions which are capable to fulfill these requirements. We have introduced a mechanism which enables existing signature verification services to integrate external services supporting the verification of national and proprietary signature formats. This facilitates the dynamic enhancement of supported signature formats. In addition, we have highlighted the need for previous data validation and have presented a solution on validating meta and document data. Although we have presented our solutions on the basis of the OCD container format and its software modules, they are also applicable for several other use cases where electronic documents must be exchanged and processed. Anyhow, our presented approach contributes to more efficient, time saving, and cost reducing e-Government and e-Business applications.

Additionally, we have elaborated possibilities to make verification and validation services available via cloud computing. The movement of these services to the cloud allows for additional cost savings and enables higher scalability. The incorporation of encrypted OCD containers and documents in the presented approach as well as the definition of interoperable document data types are subjects to be addressed in our future work. This might also help bypassing security and privacy concerns with respect to cloud computing.

REFERENCES

Catteddu, D., and Hogben, G. (2009). *Cloud Computing - Benefits, risks and recommendations for information security*. ENISA

Cloud Security Alliance. (2011). SECURITY GUIDANCE FOR CRITICAL AREAS OF FOCUS IN CLOUD COMPUTING V3.0.

Director General of the Lithuanian Archives Department, 2009. *Specification ADOC-V1.0 of the electronic document signed by the electronic signature*. Valstybės žinios 108-4574, https://signa.mitsoft.lt/static/signa-web/webResources/docs/ADOC_specification_approved20090907_EN.pdf

ETSI, 2009. *ETSI TS 101 903, XML Advanced Electronic Signatures (XAdES)*. Version 1.4.1.

ETSI, 2010. *ETSI TS 102 778-3. Electronic Signatures and Infrastructures (ESI); PDF Advanced Electronic Signature Profiles; Part 3: PAdES Enhanced - PAdES-BES and PAdES-EPES Profiles*. Version 1.2.1.

ETSI, 2012. *ETSI TS 102 918, Electronic Signatures and Infrastructures (ESI); Associated Signature Containers (ASiC)*. Version 1.2.1.

European Commission, 2011. *European Commission Decision on Establishing minimum requirements for the cross-border processing of documents signed electronically by competent authorities under Directive 2006/123/EC of the European Parliament and of the Council on services in the internal market, notified under document C(2011) 1081. 2011/130/EU*. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:053:0066:0072:EN:PDF.

European Union, 2006. *Directive 2006/123/EC of the European Parliament and of the Council of 12 December 2006 on services in the internal market*. Official Journal of the European Union. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:376:0036:0068:en:PDF.

Mell, P., and Grance, T. (2010). The NIST definition of cloud computing. NIST. http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf

Leitold H. et al, 2009, *Mediabreak resistant eSignatures in eGovernment - An Austrian experience*. Emerging Challenges for Security, Privacy, and Trust - 24th IFIP SEC. Volume IFIP AICT 297 of IFIP Advances in Information and Communication Technologies, pp. 109-118.

Rössler T., Tauber A., 2010. *The SPOCS Interoperability Framework: Interoperability of eDocuments and eDelivery Systems taken as Example*. ISSE 2010 Securing Electronic Business Processes, pp. 122-130.

SPOCS, 2011. *SPOCS Deliverable D2.2 - Standard Document and Validation Common Specifications*. Version 1.4.0. http://joinup.ec.europa.eu/site/spocs/eDocuments/references/D2.2_Standard_document_and_validation_common_specifications.zip

Stranacher K., Kawecki T., 2012. *Interoperable Electronic Documents*. Electronic Government and Electronic Participation - Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and IFIP ePart 2012, Kristiansand, Norway, pp. 81-88.

Zefferer T. et al, 2011, *Secure and Reliable Online-Verification of Electronic Signatures in the Digital Age*, Proceedings of the IADIS International Conference WWW/INTERNET 2011, pp. 269-276.

Zissis, D., and Lekkas, D. (2012). *Addressing cloud computing security issues*. Future Generation Computer Systems, 28(3), 583–592.