ARTICLE:

Security in digital data preservation

By **Franco Ruggieri**

Introduction

For a long enough time now, all countries have been discussing moving from a paper world to a digital one. The process is turning out to be not so fast, although eventually we humans should succeed. It is to be remarked that digital 'objects', i.e. not only what is generally recognised as 'documents' such as Word or pdf files, but also images, outputs created by Computer Aided Design programs, spreadsheets, music, and the like, have a requirement in common with their paper relatives: that is 'preservation', with the, not so slight, difference that preservation is entirely different in the two cases. One purpose between them is common, though: making it possible to retrieve and to 'read' preserved objects. A problem is the length of time that digital data can be preserved. It might be necessary to preserve digital data for decades,[1] and maybe even longer: for example, notarial deeds are supposed to last forever. Preserving documents on paper seems very easy, but it is incredibly difficult to create back-ups, to use the Information Technology term, even more if these back-ups are located far from the 'central' preservation site, especially in case of 'unique' documents, such as where there is no 'official' copy. Making copies of 'unique' digital documents (think of digitally signed data), in contrast, takes seconds. This makes, in fact, the term 'unique digital document' an oxymoron.[2]

There are two pillars relating to preserving digital objects:

1. Ensuring it will be possible to later retrieve them, even amongst billons of other digital objects, and to 'read' them.

2. Ensuring they will not 'disappear', that is: they are 'securely preserved'.

The retrieval of digital data is covered by a number of ISO standards, the paterfamilias of which is *ISO 14721– Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model*, originally produced by CCSDS (Consultative Committee for Space Data System) of the National Aeronautics and Space Administration as a recommendation, and given the status of an ISO Standard in 2003. The current version is dated 2012. It 'addresses a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives.' Moreover, for each of the areas, it stresses the need to take care of the security of the preserved objects in order to prevent them from being deleted or modified or becoming illegible for any reason. However, it provides no security related measure.

On the one hand, ISO 14721 strengthened, with its specifications, some of the previously existing standards, a few of which have been subsequently upgraded following the ISO 14721 path, and on the other hand it gave way to an ever increasing number of specifications, standards and legal instruments dealing with a number of topics: from how to create digital data, the creation of digital data from analogue documents (paper or other analogue media), up to how to set up metadata suitable to enable the retrieval of preserved objects.[3] It is worth remarking

[1] Domestic legislation often sets out period of time that document have to be retained for certain categories of document, such as accounting records, and industries have different requirements for other types of document, depending on the nature of the work they undertake.

[2] An interesting and poetic digression on this topic can be found in an article by Stephen Mason, 'Electronic evidence and the meaning of "original"', *Amicus Curiae* The Journal of the Society for Advanced Legal Studies, Issue 79, Autumn 2009, 26 – 28, available at http://sas-space.sas.ac.uk/2565/1/Amicus79_Mason.pdf

[3] Among the umpteen documents and standards that now exist, it is worth reminding the reader of at least the following: ISO 15489-1:2001 – Information and documentation -- Records management -- Part 1: General; ISO/TR 15489-2:2001 – Information and documentation -- Records management -- Part 2: Guidelines; ISO 23081-1:2006 – Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles;

that this number of archival related standards is still increasing.

Unfortunately none of these documents addresses in depth the other important topic: security in preserving digital data. Many of these standards and specifications recommend the reader to refer to ISO/IEC 27000 family of documents. Other ones, such as ISO/TR 15489-2, consider security in more depth, but not in enough detail to provide users with sufficient guidance. To give the reader an example, clause 4.2.5.2 of ISO/TR 15489-2 reads:

> 'Development of appropriate categories of access rights and restrictions is based on the organization's regulatory framework analysis, business activity analysis and risk assessment. Reasonable security and access will depend on both the nature and size of the organization, as well as the content and value of the information requiring security.'

Further on, clause 4.3.7.1 reads:

> 'It is important to determine efficient and effective means of maintaining, handling and storing records before the records are created and then to reassess storage arrangements as the records' requirements change. It is also important that storage choices be integrated with the overall records management programme.'

Similarly, other clauses provide more detailed recommendations, but readers can find just 'what' is to be done, not 'how' to do it in order to achieve preservation security.

A change occurred in April 2012, when the European Telecommunications Standards Institute (ETSI), one of the three European standardisations bodies officially recognised by the European Union Commission, published two documents marked '101 533'.[4] These 101 533 documents are:

> ETSI TS 101 533-1 v1.2.1 (2011-12) – Electronic Signatures and Infrastructures (ESI); Data Preservation Systems Security; Part 1: Requirements for Implementation and Management
>
> ETSI TR 101 533-2 v1.2.1 (2011-12) – Electronic Signatures and Infrastructures (ESI); Data Preservation Systems Security; Part 2: Guidelines for Assessors

Before considering both these documents, we will first consider the most recent ISO standard related to digital archival: *ISO 14641-1:2012 Electronic archiving -- Part 1: Specifications concerning the design and the operation of an information system for electronic information preservation (January 2012)*.

ISO 14641-1

This ISO standard, derived from the French norm AFNOR *NF Z42-013*,[5] is declared as the first part of a family and provides what, despite the term used 'specifications', are better to be referred to as 'policies'. The norm in fact describes 'what' is to be done, not 'how' it shall or should be done. The only exception is related to converting analogue documents (paper, parchment, video, chemical photograph, vinyl record, microform, etc.) into digital ones: there are six pages dedicated to this issue out of 34 normative pages. Just to give the reader an idea of how deep this ISO 14641-1 goes in providing instructions on how to implement this analogue-to-digital conversion, one sentence in clause 10.2.3.3 '*Preparation of microform documents',* reminds the operator that *'Microform documents shall, if necessary, be cleared of dust before digitization.*'

On the other hand, ISO 14641-1 makes a very important assertion with the following statement that is included in clause 13.2.1 'Service contract', where item j) reads 'insurance policies contracted by the third party covering any activity-related damages'. This addresses a very basic issue: preservation jointly with 'archival'[6] services must be provided by

ISO 23081-2:2009 – Information and documentation -- Managing metadata for records -- Part 2: Conceptual and implementation issues; ISO/TR 23081-3:2011 – Information and documentation -- Managing metadata for records -- Part 3: Self-assessment method; ISO 30300:2011 – Information and documentation -- Management systems for records -- Fundamentals and vocabulary.

[4] Available in electronic format from http://pda.etsi.org/pda/queryform.asp.

[5] Available at http://www.boutique.afnor.org/norme/nf-z42-013/archivage-electronique-specifications-relatives-a-la-conception-et-a-l-exploitation-de-systemes-informatiques-en-vue-d-assurer/article/773362/fa125098 .

[6] It is worth mentioning the original distinction between 'archival' that addresses, in the main, the purpose of being able to retrieve archived digital objects (by means of metadata), assuming that their preservation is somehow assured, and 'preservation' that focusses on providing security measures suitable to ensure that the binary content of the preserved digital objects is not tampered with. The border between these two terms is somewhat reduced, since one cannot provide archival services without applying security measures suitable to ensure the digital objects integrity and, therefore, 'persistence'.

organisations that are financially robust enough, even to the extent of bolstering this financial strength with an insurance policy. To further streamline the matter, Italian Legislative Decree No 82 of 2005 (amended a number of times in order to keep it up to date), at article 44-bis (specific to 'accredited' digital preservation services), requires at paragraph 3 that accredited preservation providers be legal persons with at least 200.000,00 euros capital. The term 'accredited' in this legislative instrument means 'recognised as fit for this type of services' by Agenzia per l'Italia Digitale, an Italian governmental body that over time has changed its name several times: from AIPA (Agenzia per l'Informatica nella Pubblica Amministrazione), to CNIPA (Centro Nazionale per l'Informatica nella Pubblica Amministrazione), to DigitPA (just a name) and currently, but presumably not 'finally', to AgID.

Very little is said in ISO 14641-1 about security and in particular on 'information systems security'. Clause 5.5 headed 'Security', provides some detail in sub-clause 5.5.1, headed 'Management and organization of security'. The clause reads: 'NOTE – For security requirements, reference should be made to ISO/IEC 27001 and associated standards.' There are further 13 more lines that, after having said 'The management system for security shall be distinct and separate from the administration of information system operations or telecommunications systems', lists 8 principles to be met to ensure security, among which is: 'management of the keys of premises, security systems for detection, intrusion and alarms; compliance of hardware with regulations concerning human safety'. And that is all on security. Similarly, risk assessment is dealt with in 16 lines in clause 5.5.2. Let us go through the gist of these 16 lines. After having recommended that preservation services must be proactive and not reactive, by drafting security policies, it states: 'The organization shall undertake an information security risk analysis, and document the results obtained.' The focus then moves on to dealing with storage media, both live and backup, by analysing 'vulnerability risk factors' that are customised and based on the different types of media. After having done this analysis, the outcomes should be reflected in the security measures, keeping in mind a balance between 'costs of implementation, security achieved and risk evaluation', and, consequently, on the security procedures.

However, clause '1. Scope' prevents any misunderstanding: 'This part of ISO 14641 is not applicable to information systems in which users have the ability to substitute or alter documents after capture.' In other words, it up to the archiving organisation to define by itself a set of measures that are suitable to prevent the substitution or the alteration of documents and to ensure that they can be retrieved and legible, that is to counter format and media obsolescence.

This is not a criticism of ISO 14641-1. The point is that, unfortunately, not even this ISO standard provides digital preservation operators with sufficient guidelines on security. The simple pointer to *ISO/IEC 27001 – Information security management*, that can be found in the previously mentioned NOTE of clause 5.5.1: 'For security requirements, reference should be made to ISO/IEC 27001 and associated standards', does not help much: arguably, the digital preservation domain is so peculiar that it needs a thorough customisation of ISO/IEC 27001-27002 with specific provisions and guidance. It is this reason that led ETSI to draft the two 101 533 documents.

The ETSI 101 533 family

The history

The author of this article proposed the development of this type of specification in 2009. The idea was submitted to UNINFO, which is a standardisation body federated to the Italian official standardisation body UNI (Ente Italiano di Normazione). A UNINFO working group was launched on 12 June 2009 under the heading 'security in digital (document) preservation'. This led ETSI to make an application to the European Union to fund a similar effort that was to be performed by ETSI itself. The EU Commission agreed, and funded an ETSI work group that had its first meeting on 15 June 2010. The author acted as the coordinator of both work groups, which made it very easy to ensure they cooperated. Eventually, in the last week of February 2011, both groups finalised the content of the specifications. They were, obviously, in English, and so ETSI could claim ownership of the copyright, by quickly publishing them. ETSI subsequently permitted UNI to translate them into Italian. This translation was performed between 2011 and 2012, during which a number of amendments were identified and submitted to ETSI. Finally in April 2012 both ETSI, in English, and UNI, in Italian, published the currently available documents. It is to be remarked that these documents are based on, and refer to, ISO/IEC 27002:2005, so it is necessary to jointly read the ETSI documents and ISO/IEC 27002:2005.

Unfortunately, Murphy's law[7] always prevails, so in 2013, a few months after both ETSI and UNI published their sets of documents based on version 2005 of ISO/IEC 27001 and 27002, ISO/IEC issued a new version of these very standards: ISO/IEC 27001:2013 and ISO/IEC 27002:2013. Fortunately, the differences are not dramatic, so it is not necessary to rush to update the ETSI and UNI documents. Therefore, the ETSI documents are perfectly valid, provided that the reader, with just a little extra effort, keeps an eye also on the 2013 versions of ISO/IEC 27001 and ISO/IEC 27002 when implementing the ETSI documents.

These documents were funded by the European Union, which implied that they were meant to be applicable throughout the Europe Union. Thus, depositors, or even preservation providers wishing to outsource some of their services, may even request foreign service providers to demonstrate that they are capable of abiding by this specification before entrusting them their data objects. They can demonstrate this by exhibiting either a suitable assessment by a 'conformity assessment body', in accordance with the recent EU Regulation 910/2014,[8] or, more preferably, the 'qualified trusted service provider' status awarded by the EUMS relevant governmental body.

A clarification is necessary, though, to prevent misunderstandings. The preservation service, referred to in the ETSI documents, should be performed by skilled organisations, whether in-house or external providers. Therefore, it is to be provided by a specific 'body', be it one specific division of a company or an external service provider. This depends on two factors:

1. Preservation requires a set of measures that a duly skilled body is capable of implementing.
2. If the preservation of digital objects is concentrated in one body, individual employees should not bear a responsibility for preserving the data. If (absurdly) one company requires each employee to preserve all of their documents without a form of central control, it might be impossible to find any document in the long run especially if it is a large, possibly multinational, company.

One central preservation site is highly desirable, because if the preservation of digital objects is delegated to each employee, there will be two risks: security measures would probably be applied unequally by the various employees, and retrieving documents scattered among several persons would be difficult – perhaps impossible.

The specifications

ETSI TS 101 533-1

As the title states, this document specifies '*Requirements for Implementation and Management*', hence the document type 'TS', that is, 'Technical Specification'. This Technical Specification structure mirrors that of a previous ETSI document: *ETSI TS 102 573, v2.1.1 – Electronic Signatures and Infrastructures (ESI); Policy requirements for trust service providers signing and/or storing data objects*. ETSI TS 101 533-1 provides provisions on how to securely implement and manage a preservation system of digital objects, any kind of digital objects: documents (i.e. file types readable with Microsoft Office Word or Excel, or pdf files), any kind of image files, movie files, Computer Aided Design (CAD) files, etc. Additionally, where the preservation of digital objects is performed via cloud based solutions, the ETSI documents also apply, since the prescribed security measures are to be implemented on any system used in the cloud. Security, and in particular 'preservation security' is not a trivial issue that should, no: 'must' be duly implemented.

Two service types are addressed:

1. The Core Services, which are mandatory. These services provide depositors with the basic service of preserving what was deposited in an unchanged way for the period of time agreed upon. No additional service is meant to be performed. If the documents do not meet relevant legal rules or any other rule relating to the term of retention, or if the document format becomes obsolete and can no longer be read, the responsibility is not

[7] The Oxford English Dictionary (electronic version) provides a definition at 3.3: 'Murphy's law: a name humorously given to various aphoristic expressions of the apparent perverseness and unreasonableness of things (originating from the U.S.)'. In practice it means: 'Anything that can go wrong will go wrong'.

[8] 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, OJ L 257, 28.08.2014, p. 73–114.

upon the Data Preservation Service Provider (DPSP).

2. Optional services, called 'Extended Services', that a DPSP may want to provide to its customers. Not all of the possible services of this kind have been specified in the documents: it is simply impossible, since there is no limit to the requirements of an individual customer. Not all of these services depend on the customers' requirements, although the Italian legislation makes some of them mandatory.

Three kinds of provisions are specified for both service types:

'must' – provisions marked with such verbal form are mandatory: they can be superseded only by 'force majeure', for example the applicable legislation;

'recommended/should' – this form has a specific meaning: provisions marked with it can be disregarded only after having carefully assessed in depth the possible consequences of the decision to not abide by the recommendation; this assessment and the deriving justifications must be documented and submitted to auditors (be they internal or external) who, when performing an assessment or an audit on a DPSP, are required to evaluate if these justifications suffice; auditors may even refuse to perform an audit, should they deem the explanations as unsatisfactory;

'may/optional' – the implementation of such provisions is for the DPSP to decide.

The Data Preservation Service is very specific. This led ETSI and UNINFO to decide that it was necessary to 'customize' ISO/IEC 27002:2005, which is what is called a 'general purpose' standard. As a consequence, while ETSI TS 101 533-1 assumes that all ISO/IEC 27002:2005 measures apply that are by default 'recommended', in this Technical Specification, a number of them are deemed to be 'mandatory', others are deemed as 'optional', and yet other measures are deemed as 'not applicable'. Additionally, a large number of new measures are added that are considered as mandatory, recommended or optional.

A few specific measures relating to the Data Preservation Service Provider

As a first example of these measures, albeit this may seem trivial, is a basic requirement on DPSPs to ensure that their decisions are independent from their service providers' or customers' will. Even if a DPSP is a department of one company, its decisions must be free from undue interference by other departments or senior managers. The following example illustrates what must not occur. The related preservation logs are to be time stamped by a Time Stamping Authority (TSA), so it is impossible to tamper with these data without it being evident. Imagine that a DPSP discovers that a significant mistake occurred a few years earlier. In such circumstances, it is impossible to rectify the mistake. But if there is an improper relationship between a DPSP and the TSA, the DPSP might seek to amend the mistake by producing a new correct version of the data at issue, by submitting them to the TSA and by telling the TSA to 'stop working for a few minutes, change your clock and calendar to this time and date I am telling you, then apply a time stamp on these new data I am submitting you. After having done this you can reset your calendar and clock'. Obviously this would be unacceptable.

Another example is that it is necessary to implement the separation of roles – this is a normal basic security requirement in order to prevent misdeeds. Consideration is not given to this elementary point – rather the spotlight will focus on two aspects that DPSPs must be careful to be aware of.

1. The development, test and operational environments must be separated. While separation between the development and operations environments is a truism, specific attention must be paid on separating the test and operational environments. In fact, the final acceptance of an application, of a system or of one of its HW [hardware] components, is more often than not based on a test run on the operational data set. This is dangerous for at least two reasons: the first, and obvious one, is that nobody can be sure that the test will not alter the data set. The second reason is that during these tests, the input and output will be carefully looked through, and consequently a data privacy violation may occur. As a consequence, before performing these tests, the operational data set must be

copied in the test environment and, when performing this migration, personal data must be made anonymous.

2. The second aspect, trivial though it may appear, is that persons that have developed HW and/or SW [software] components [hardware and software components of a system] and submit them to acceptance tests, must not participate in such tests in a way that enables them to perform undue acts, such as altering the tested component behaviour so to make the tests appear as successful. To repeat, this may appear as trivial, but we have to be sure this never happens.

A third example follows. Apart from the usual differentiation of access rights among system operators, system administrators, data owners, auditors and authorities, and the provision of an effective access control system to rooms that host computing systems storing and operating on the preserved data, another caution was specified: sensitive operation officers should be given, where necessary, credentials to be used when ‘under duress’. These credentials would let these officers authenticate when forced to do so by non-authorised persons, but would activate a ‘forced intrusion’ alarm and direct the operations performed by the officer into a ‘sandbox’ where whatever action is carried out does not affect the real data, but just a simulated set of them.

Another example (as previously hinted at) that distinguishes Digital Preservation Systems from paper archives, is the possibility to easily set up back up sites and disaster recovery sites, even remote ones, in order to provide for business continuity. As a consequence, TS 101 533-1 requires DPSPs to set up remote back up sites and to keep them updated, according to a specific plan, and to set up a Disaster Recovery Team, which should be made promptly available in case of disaster.

Finally, it is interesting to remark that attention is given to a particular kind of malware that was called ‘Presentation Corruption Agent’ (PCA), that is ‘macros, hidden executable code, hidden or difficult to detect worksheet formulas and hidden byte sequences that are ignored by the originally intended presenting application but that can be recognised when the data object is processed by different applications’. A PCA can actually change a document presentation without affecting the document binary content (macros, formulas, etc. are included in the document), so it cannot be countered by the use of advanced electronic signatures. PCAs are therefore particularly insidious. Measures against PCAs are obviously ‘Extended Services’ and are described in clause 6.3.5: (i) document formats should not be among those indicated by standardisation bodies as suitable to host PCA, and (ii) if a DPSP provides this kind of Extended Service, it shall have in force suitable procedures to verify if deposited documents are free of PCA and shall be able to demonstrate these procedures effectiveness. The latter is easier said than done.

In any case the process cannot go beyond the preservation of the data and affect how a digital object is printed or displayed, for instance. An example would be a colour photograph. The data comprising the photograph might have preserved the image correctly. If the photograph is printed, the image might, for instance, be printed in grayscale, and not colour, because of the setting on the printer. That the photograph has not been printed in colour will not affect the preserved image.

ETSI TR 101 533-2

This document is addressed to assessors, be they internal or external, and, as such, there are neither ‘must’ nor ‘shall’ provisions: a certified assessor should know perfectly well how to conduct an assessment. The purpose of this document was only to provide assessors with guidance on how to deal with assessing this peculiar entity called DPSP. This document is structured as a mirror of its sister document, TS 101 533-1, so that assessors can easily establish the requirements that they should ascertain. Assessors will also have to be aware of ISO/IEC 27002:2005 and of the previously mentioned ETSI TS 102 573, in order to be aware of what they are required to verify, which includes such items as:

1. The periodic Risk Assessment outcomes are promptly reflected in the DPSP Security Policies and Procedures.

2. The organisation chart of the company to which the DPSP belongs proves the independence of DPSP decisions from any undue interference.

3. The team the DPSP must set up to face disasters in order to have a suitable Disaster Recovery Plan will operate as such only in case of actual disasters or in drills.

4. The Disaster Recovery Plan is duly drafted and, mainly, tested with periodic drills.

Conclusions on the ETSI documents

It is interesting to observe that, although the ETSI documents are based on the 2005 version of ISO/IEC 27001/27002, should a DPSP have a certification under ISO/IEC 27001:2013, the provisions in ETSI TS 101 533-1 still apply. The significant point is that the DPSP must implement the measures specified in such ETSI document, taking into account the new 2013 ISO/IEC 27002 version. There could be some measures that did not exist in the 2005 version, the implementation of which should be assessed by the DPSP itself, without a TS 101 533-1 'supporting opinion'.

Italian technical and legal provisions

It has been noted above that UNINFO and ETSI have worked together, and that ETSI acquired the copyright on the '101 533' by being the first to publish them. It has also been noted that UNINFO/UNI was authorized to translate them into Italian. What was not mentioned is that UNI/UNINFO was authorized by ETSI to develop and publish a third document, containing a number of Italian-specific provisions for the purposes of the Italian legal and customary 'world'. The purpose is to integrate the ETSI TS 101 533-01 and ETSI TR 101 533-02 into the Italian legal and customary environment. A few of these addenda requirements are indicated as mandatory, even for assessors, among which is the obligation to provide a DPSP with the result of assessments conducted on its suppliers of services related to the provision of Digital Preservation Services.

From the legislation viewpoint, it is interesting to remark that in Italy a Decree by the President of the Council of Ministers (DPCM) issued on 3/12/2013, published in the Italian Official Journal on 12/3/2014, is entirely dedicated to digital document preservation, and that both ETSI TS 101 533-1 and ETSI TR 101 533-02 are among the standards and specifications listed in the decree as suitable to meet the requirements of the Decree. In other words, in Italy, if a DPSP abides by these ETSI Specifications, it is deemed as conforming to the DPCM requirements in order to achieve accreditation by the specific governmental body, the Agenzia per l'Italia Digitale. It is possible to argue that this accreditation is a minor issue, but it is to be taken into account that public administrations can entrust the preservation of their digital documents to a private DPSP only if the latter is accredited. For this reason, accreditation is of particular importance to DPSPs and, therefore, ETSI TS 101 533-1 is important too.

Finally, on 10 April 2014, Agenzia per l'Italia Digitale issued what is called a 'Circular letter' specifying the methods with which preservation service providers can be accredited by the Agenzia, and how the Agenzia must behave in accrediting and inspecting such providers. Because this 'Circular letter' is so recent, there is no finalised accreditation yet, but a number of preservation service providers are at work to set up their security measures consistently with all the relevant requirements, including ETSI TS 101 533-01. The process is on!

Franco Ruggieri is an independent consultant on ICT security, e-signatures, e-invoicing, digital preservation, and registered e-mail. He contributed to the drafting of a number of ETSI specifications between 2001 and 2011, including ETSI TS 101 533-01 and ETSI TR 101 533-02, acting as the leader of the relevant task force.