InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Italy

Title: General Study 05b – Guidelines and Recommendations for E-Mail Records Management and Long-Term Preservation

Final Report

Status:	Final (public)
Version:	2.3
Date Submitted:	May 2011
Last Revised:	February 2016
Author:	The InterPARES 3 Project, TEAM Italy
Writer(s):	Massimiliano Grandi Researcher, TEAM Italy
Project Unit:	Research
URL:	http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05b_final_report.pdf

Document Control

Version history			
Version	Date	By	Version notes
2.0	2011-05-30	M. Grandi	First draft.
2.1	2011-06-01	A. Allen	Minor copy edits.
2.2	2011-06-02	M. Grandi	Minor copy edits.
2.3	2016-02-29	R. Preston	Copy and minor content edits for public version.

Table of Contents

1. CURRENT STATE OF THE ART

1.1 BACKGROUND

According to estimates by The Radicati Group,[1] in 2009 approximately 280.2 billion e-mails were sent each day. Even if spam accounts for four fifths of all e-mails sent,[2] the remaining part is still a staggering number. Moreover, many organizations have to develop strategies to deal with both legitimate and spam messages.[3] These figures alone clarify the role of e-mails in our society: they are by far the most widespread and ubiquitous system in the world to exchange digital documents and the use of them is constantly on the rise. The largest problem for archivists and records managers is to gain intellectual control over this huge mass of material and enable recordkeeping systems and digital repositories to cope with it without being disrupted by its sheer amount.

E-mails are not per se records and do not constitute a series. As the Division of Archives and Records Management of the State of Washington rightly states:

> E-mail is a means of sending or receiving information, not a type of record. Information generated or received on an e-mail system needs to be managed according to the informational content of the message[4]

Although it is often understood as a type of document, e-mail is a term that indicates a system by means of which many different information objects are carried. Nevertheless, by now the word 'e-mail' is also a metonym for the digital objects that are transported. Just as the expression 'correspondence' can stand for a large gamut of documents, some of which hardly could be considered records, such as fliers or complimentary press copies, 'e-mail' can indicate a wide range of entities. E-mail is a technology that has been developing since 1971,[5] based on standards and protocols that are the foundation of a "method of exchanging messages on the Internet."[6] Of course, many e-mail messages are also records, in the same way that some correspondence can be records. When an e-mail is a record, we may complement the above-mentioned definition of the Division of Archives and Records Management of the State of

---

[1] The Radicati Group Inc. *Microsoft Exchange Market Share Statistics, 2005* (Palo Alto, CA: The Radicati Group July 2005), 13. Cf. download.microsoft.com/download/E/8/A/E8A154BF-CC35-4340-BD26-6265CDB06B6E/ExStats.doc (last accessed March 16, 2010).
[2] Ibid., 14.
[3] For example, in 2002 the 14,000 employees of the U.S. Department of Energy had to process about one million e-mail messages a day, cf. A. Roberts. *Blacked Out. Government Secrecy in the Information Age* (Cambridge, UK: Cambridge University Press 2006), 213.
[4] Agencies of Washington State Government, Office of the Secretary of State Division of Archives and Records Management. *General Records Retention Schedules* (Olympia, WA: Washington State Records Committee 2005), 6. http://library.evergreen.edu/recordsmanagement/RecordsSchedules/WAGSRet.pdf (last accessed June 2, 2011).
[5] G. Pontevolpe and S. Salza. *General Study 05 – Keeping and Preserving*, version 4.0 (The InterPARES 3 Project, TEAM Italy, June 2009), 4. http://www.interpares.org/rws/display_file.cfm?doc=ip3_italy_gs05_draft_report_v4_RESTRICTED.pdf (restricted draft, i.e., access to the file is restricted to the InterPARES participants - last accessed June 2, 2011).
[6] Ibid., 7.

Washington[7] by adding that the e-mail is also to be handled according to the business transaction to which the record relates. Any e-mail is composed of a message header, containing several elements of information related to the process of dispatch and delivery, followed by a body (which typically contains the information that the sender wants to transmit – although in theory the body of the e-mail could be empty. A body can be formed of one or more parts: when there is only one part, the entity is commonly referred to as a "simple e-mail", whereas if the e-mail contains multiple parts, it can form many different combinations. This makes e-mail an extremely powerful and flexible application, through which it is possible to convey a great variety of digital entities encoded in different formats. Probably the most known and widespread kind of multipart e-mail is a message formed of a plain text section visualized just below the header section and linked with one or more attachments. For records management and archival science purposes, it is particularly important that a single e-mail can carry one or more attachments that from the perspective of diplomatics, could be regarded as parts of a single record, attachments of a record or even separate records. Diplomatics makes it possible to envisage different sorts of manifestations: for instance, one of the attachments of an e-mail could be a record, while the other ones and the plain text section, containing the links to all the attachments, may be considered according to the diplomatic analysis just the attachments of this record.

1.2 KEY ISSUES AND CHALLENGES

The remarks at the end of section 1.1 show that the intellectual organization of the elements of an e-mail message can be completely different from the physical organization, which is a serious problem for records managers and archivists.

E-mail objects, thanks to the use of the MIME standard,[8] can be formed by a huge number of components (among which we can include digital signatures, threads of messages, files encoded in any format, alternative versions of the same content) and associated with any kind of information (e.g., links, images, mark-up tags, distribution lists). Such e-mails are complex digital structures. Physical complexity, in turn, may involve diplomatic complexity. An appropriate intellectual and physical treatment of e-mails in a recordkeeping system must be grounded in the correct identification and interpretation of the elements of form and digital components which constitute each entity that is to be included into the system. Moreover, the presence of various components and all sorts of formats call for adequate actions to assure the long-term preservation of the e-mail records selected for permanent retention.

It may become difficult to correctly determine important metadata elements, such as the persons involved in the creation of an e-mail record or the types of context characterizing the

[7] Cf. 1.
[8] MIME stands for "Multipurpose Internet Mail Extensions." For a brief overview of the MIME standard, cf. G. Pontevolpe and S. Salza, 2009, pp. 12-13, 16.

action the e-mail record is intended to support,[9] in consideration of the extreme variability of manifestations which characterize e-mails. E-mail technology has enabled users to avail themselves of many features typical of word processors in a simpler environment that may run in a wide array of different platforms. Pliability and pervasiveness, well supported and further strengthened by the MIME specifications and all the Internet protocols, are able to originate a great number of documentary forms that may hide from a user the crucial pieces of information needed to properly describe an e-mail record. The following examples illustrate this point: a reply to an e-mail can be composed by inserting sections of text amidst parts of the former e-mail, which may lie inside the new e-mail message without any metadata or visible mark revealing their presence at a glance; it is possible to merge multiple e-mails into one message; in theory an e-mail can be sent from an address which has been set up solely to deliver it; or a sender can hide the identity of the addressees of a message by means of a distribution list. If on the one hand such characteristics make e-mails flexible instruments that facilitate communication, on the other hand a strong intellectual control over this kaleidoscopic array of documentary forms is needed for the purpose of appropriately identifying all the metadata elements of an e-mail record captured in a recordkeeping system.

The problems arising from the potential complexity of the documentary form of e-mails are compounded by the huge number of these objects that typically pass through the IT system of any organization, so that the overall daily workload entailed by the processing of e-mails is extremely time-consuming for the staff and can seriously interfere with the ordinary activities of an organization. Difficulties which arise out of not knowing how to properly treat e-mails can easily result in bad document and records management practices, all the more so as e-mails are often created and received in IT environments that can isolate them from all the other records an organization produces in the course of its business operations. In fact, e-mails are normally managed by means of dedicated applications, such as e-mail clients and webmail programs,[10] which do not communicate with other systems, in absence of specific measures aimed at this purpose. E-mail clients can be set up in any workstation and are usually proprietary tools which download and store e-mails in the hardware resource where they are located, so that they may often remove e-mails from the centralized corporate server. On the contrary, webmail applications, as their name shows, may be exclusively used through a web interface and, as often in the case of e-mail clients, are ordinarily subject to restricted access, reserved for those who are able to authenticate their own identity by means of a set of parameters (normally a user name and a password). Moreover, both solutions often use proprietary features and formats, so that they cannot be integrated in the IT network of an organization without proper action.

To worsen the situation there is also the high average number of accounts that can send and receive e-mails in the IT system of an organization. In fact, many organizations set up

[9] Cf. InterPARES 3 Project, TEAM Canada. "Intellectual Framework Version 2.0," September 2008, 4-6. http://www.interpares.org/display_file.cfm?doc=ip3_intellectual_framework.pdf (last accessed March 19, 2010); InterPARES 3 Project. "Template for Diplomatic Analysis," 1-4. http://www.interpares.org/display_file.cfm?doc=ip3_template_for_diplomatic_analysis.pdf (last accessed March 19, 2010).
[10] G. Pontevolpe and S. Salza, 2009, p. 8.

individual business accounts for each of their employees or co-operators, which at least partially state their names in the local part of the addresses. Such personal accounts are added to those that correspond to the offices and functional roles of an organization (e.g., the e-mail address of the human resource management department or the library assistant manager) and can be in a sense regarded as the "institutional" set of e-mail accounts devoted to handling all the tasks related to its specific mandate. Nonetheless, in almost every work environment it is a common habit to carry out corporate duties by means of individual business accounts and such a use is frequently not only accepted, but even formally acknowledged in regulations and policies.

This usage entails serious consequences from a records management perspective. First, possible access points to the IT system multiply and there may be hundreds of them within a single organization. This makes it far harder to ensure that e-mails which are records will be captured in the recordkeeping system. But the most dangerous consequence of such a practice is that business and private e-mails end up mixed together and sometimes even individual messages contain information concerning both personal matters and corporate processes. Although many organizations have enacted policies that forbid employees from combining personal (non-business) content with data relevant to the work environment in the same document, the enforcement of these provisions is problematic. Some judges have determined that employers are not entitled to inspect the e-mail business account of an employee, even if the investigation was just aimed to ascertain the use of the account by the employee in relation to professional activities.[11] On most occasions, however, judges have decided that such checks performed by employers are legitimate.[12] The whole issue is still a moot point for jurisprudence. At the minimum, an organization has the responsibility to draft a well-conceived and precise policy to face with some confidence possible legal action filed by a worker who deems that the organization has violated his or her privacy by such an inspection.[13]

---

[11] Cf. M .S. Hornug. "Think Before You Type: A Look at Email Privacy in the Work Place," *Fordham Journal of Corporate & Financial Law* 11, no. 1 (2005): 144-145. http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1205&context=jcfl (last accessed May 15, 2011).

[12] Ibid., pp. 140-144.

[13] A comprehensive review of the acts and legal decisions worldwide that relate to this subject is beyond the scope of this paper. It is to be noted that the Italian Data Privacy Agency issued in 2007 a deliberation concerning guidelines for the use of e-mails and the Internet in working environments, where, with reference to the checks an employer may carry out in a workplace on the electronic communications of his or her employees, it is stated that "Controls are only lawful if the relevance and non-excessiveness principles are complied with."; cf. Italy, *Provvedimento 1 marzo 2007, Trattamento di dati personali relativo all'utilizzo di strumenti elettronici da parte dei lavoratori* (*Deliberation March 1, 2007 - Guidelines Applying to the Use of E-Mails and the Internet in the Employment Context*), § 6.1, DPA Commissioner, *Gazzetta Ufficiale* 58 (March 10, 2007): 7. An English version of the deliberation is available at http://www.garanteprivacy.it/garante/doc.jsp?ID=1408680 (last accessed April 14, 2011). The reading of the whole deliberation shows that in Italy the establishment of proper rules that may allow for both corporate business needs and personal privacy is difficult. However, many sources agree that the question is, at a minimum, controversial. Cf., for USA, L. Rosencrance. "Sidebar: Employee Rights and Relations," *Computerworld*, 2004. Available at http://www.computerworld.com/s/article/93472/Sidebar_Employee_Rights_and_Relations?taxonomyId=018 (last accessed March 17, 2010); for Italy, M. L. Piccinni. "Privacy ed e-mail aziendale, una guida per capire" ("Privacy and Corporate E-mails, a Guide to Understand the Issue"), *SearchSecurity.it*, 2009. Available at http://searchsecurity.techtarget.it/articoli/0,1254,18_ART_104167,00.html?lw=18 (last accessed March 17, 2010); for a worldwide perspective, see Privacy International. "PHR2006 - Privacy Topics - Workplace Privacy#Email, Internet Use and Blog Monitoring," 2007. Available at http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-559090 (last accessed March 17, 2010).

Obviously, in transactions where corporate records are produced, any hypothesis of use of private personal accounts must be absolutely ruled out, in spite of some legal and political decisions that seem to support the viability of such an option.[14] Such a practice can give rise to daunting scenarios, which may involve the incapacity to guarantee the authenticity, reliability and accuracy of records, as well as other troubles connected with various infringements of rights and laws (privacy, breaches of security and so on).

1.3 CURRENT RESPONSES AND STRATEGIES

E-mail has forced all organizations to confront new substantial and complex flows of digital objects, that have to be somehow evaluated, even if only to decide to destroy them. The need of handling and appraising a gigantic number of new entities has put a significant strain on the human and technological resources organizations formerly had earmarked for records management. Therefore, beginning in the late 1990s, different strategies have been worked out to deal with the management of e-mails.

A first response has been to draw up new records management policies or to update the existing ones for the purpose of taking e-mail into consideration. For example, one might look at *RLG DigiNews*, the online newsletter of Research Libraries Group, which published in 2006 a list of policies enacted by each U.S. state.[15] The list shows that most of the states (e.g., Alabama, South Carolina, Texas) drafted documents exclusively dedicated to e-mail, while others (e.g., Oklahoma, Idaho, Washington), far more conveniently, prepared or modified existing comprehensive records management policies where e-mail is regarded as just one of the many systemic factors to be analyzed. Regardless of the discussion about the single solutions and approaches that have been proposed, e-mail records are a kind of digital record and the final goal of any provision or measure designed for them has to be their integration in a recordkeeping system. Therefore, to create documents that only contain rules and recommendations for e-mail treatment is surely a bad strategy, unless the document is functionally linked to a wider set of instructions detailing an overall consistent records management policy. In this respect, the ideal solution is to write a chapter or at least a section titled "records delivered by e-mail" in the framework of an overarching records management policy, to clarify that e-mail records are not a self-governing province in a recordkeeping system, but just records that possess specific features only with regard to the way of transmission and the structure of their format and are to be fully integrated with all the other records of the RKS.

Other initiatives have sought to deepen the understanding of the specific properties of e-mails, with regard to both their intellectual and physical form and their impact on records

[14] The Associated Press. "Sarah Palin e-mail ruling allows use of private accounts to conduct Alaska state business," Nola.com, August 13 2009. Available at http://www.nola.com/news/index.ssf/2009/08/sarah_palin_email_ruling_allow.html (last accessed March 17, 2010).

[15] R. Entlich. "You've Got Mail—Now What? Regulatory and Policy Dilemmas in Email Management. Part II. US State Environment," *RLG DigiNews*, 2006. http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3019.html and http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3021.pdf (last accessed March 18, 2010).

management workflows. Such projects are usually carried out in the framework of more general studies, which also deal with other aspects of records management. Often these projects investigate long-term preservation issues, such as DAVID,[16] an initiative developed between 2000 and 2003 in Antwerp, Belgium, or Digitale Duurzaamheid (Digital Longevity), a program established by the Dutch National Archives in 2003 and which is still ongoing. The scope and goals of these investigations differ from one another, so that to compare their respective conclusions and recommendations is somewhat difficult, but all the researchers involved have always advocated solutions aimed to fully integrate e-mail records in recordkeeping systems. In this respect, it is worth mentioning the observation of Filip Boudrez, a scholar who took part in the DAVID project: he observed that for an organization, e-mails are not only a problem to be worked out, but also an opportunity to start to build an effective records management program.[17] Since e-mail has already become the most prevalent way to exchange digital records, to neglect e-mail management automatically makes it impossible to operate any viable records management system. In this sense e-mail can be a great opportunity for records managers and archivists to demonstrate the value of their role and expertise. With respect to both the long-term preservation and the current management of e-mail records, in all the relevant literature a wide consensus exists about the benefits the use of XML[18] format can bring to an organization. Undoubtedly, apart from any other advantage connected to its extensibility and its independence from any proprietary technology, XML is a very useful tool in this field, and, in particular, XML is particularly suitable for both expressing the headers and other documentary elements of e-mails and setting up possible templates intended to create fixed configurations where the content of an e-mail can be organized. Having said that, XML must not be regarded as a panacea for any sort of problem. For example, as it concerns the long-term preservation of e-mail attachments, XML cannot directly represent a great number of different formats (e.g., certain audio files), although it may convey the metadata to help describe these formats. For procedures of migration designed for preserving such formats it is necessary to resort to other semantics, encodings and technologies. Furthermore, even if many formats can be represented in XML, it is arguable they should be transformed in XML. As long as an attachment is in an open and widely adopted format, and we have appropriate documentation and software to manage it, it does not need to be migrated to a new format: just to give an example, there is no point in transforming a txt (ascii) file in an XML file.

[16] DAVID is the acronym of Digitale archivering in Vlaamse instelligen en diensten (Digital Archiving in Flemish Institutions and Administrations).

[17] F. Boudrez, *Filing and Archiving E-Mail* (Antwerp: Expertisecentrum DAVID vzw 2006), 2.

> The second central theme of this report is the opportunity that e-mail archiving offers an organisation for putting electronic records management and recordkeeping on the agenda and into practice. The records manager or archivist can use e-mail archiving as a trigger to do something about the management and archiving of electronic documents in general.

[18] XML stands for eXtensible Markup Language. XML is a markup language created by the World Wide Web Consortium (W3C). XML is used to represent data in a standard and structured format.

1.4 A LARGER PERSPECTIVE

The erroneous belief that e-mails can be regarded as a series of records to be dealt with separately has arisen as a consequence of various factors: a seemingly simple conceptual structure of these objects, the complexity of which can be easily hidden by the similarity of some external appearances; the actual difficulties met in controlling enormous quantities of digital entities that have flooded almost every work environment; and, in the end, the ready availability of an array of tools developed on purpose for e-mails. Most of these tools have been designed with the understanding that e-mails are a particular kind of digital objects and are not meant to meet records management and archival requirements.

An analogous phenomenon occurred to a far lesser extent when faxes became widespread in the 1980s. At that time, a new way of transmitting information led some people to group together all the records conveyed by means of that technology. Nevertheless, when they were introduced, faxes were far less complicated and fewer in number than e-mails. Above all, faxes were mostly analog entities and required less effort to reengineer work processes to capture them in RMSs. However, in principle, the two situations are similar: in both cases records managers and archivists have to succeed in breaking down the intellectual constituents of the objects carried by these applications for the purpose of understanding whether they contain or are associated with records. Certainly the coming of e-mail is just one episode in which new techniques have been forcing records professionals to reanalyze the processes of creation, identification, capture, maintenance and preservation of records. E-mail, in turn, could be soon superseded by other forms of communication, such as blogs, instant messaging and interactive websites.[19] Currently, these media are not directly geared to support the transmission of records, and the information they convey is mainly informal and mostly not made up of by-products of business activities, but the state of affairs could swiftly change, as new advancements and needs could drive individuals and organizations to use them for more institutional goals.[20] It will be paramount, in this case, for records managers and archivists not to get caught unprepared and, by availing themselves of the work done to gain control over e-mails, to strive to perform a thorough intellectual analysis of the digital entities these systems transport, and to identify and properly include in recordkeeping systems also the records that may be associated with them.

[19] G. Plouin. "L'e-mail, condamné à évoluer ou à disparaître," *Le Journal du Net*. http://www.journaldunet.com/solutions/0506/050610_tribune.shtml (last accessed March 19, 2010).

[20] Actually, it is already possible to transmit records through these media, where, however, it is extremely difficult to establish the borders of documentary objects, which makes it hard to identify a record and to determine where it resides.

2. E-MAIL AS A RECORD

2.1 E-MAIL AS VIEWED BY THE INTERPARES PROJECT

E-mails are relevant to records management insofar as the messages and/or their attachments are considered to be records. Some preliminary questions must therefore be answered before dealing with methods suitable for managing e-mails in a recordkeeping system: Which characteristics must an e-mail possess to be declared a record? Which conditions must be fulfilled to ensure an appropriate integration of an e-mail in a recordkeeping system? In this respect, which problems can emerge when examining the properties and attributes of e-mails? The first phase of the InterPARES Project has identified five necessary characteristics of a digital record:

1. stable content and fixed form;
2. a link to an action;
3. an archival bond;
4. five persons (author, writer, originator, addressee, creator);
5. five types of context (juridical-administrative, provenancial, procedural, documentary, technological).

The second phase of the same project confirmed this frame of reference, but clarified – among other things – how these properties can be applied also to interactive digital objects, which can be still considered records, provided that the variability of their form and content is bounded; namely, that it is restricted by some rules and constraints that allow us to predict its behaviour[21] (i.e., as stated in the "Intellectual Framework" published in 2008 by TEAM Canada of the InterPARES 3 Project, that "the same query or interaction always generates the same result"[22]). To which extent do e-mails actually exhibit the above-mentioned features?

1. As for stable content and fixed form, e-mails are information packages that must be transmitted through a network. Sometimes they can contain some executable code, but the number of statuses and configurations such a code can create is usually finite, so there is almost always compliance with the constraint of bounded variability (at least if we leave out of consideration the number of physical manifestations that can characterize an e-mail when it is visualized by means of different clients and applications). Nevertheless, often e-mails contain links to external resources: such links are often useful also to avoid adding heavy attachments that make it difficult to dispatch and deliver them. Links can be placed both in the body of an e-mail and, if present, in its attachments. The presence of links may give rise to serious problems when the resources to which the links point are essential to use the e-mail to carry out a

[21] InterPARES 3 Project, TEAM Canada. "Intellectual Framework Version 2.0," 4-6; InterPARES 3 Project, TEAM Canada."Template for Diplomatic Analysis," 1-4.
[22] InterPARES 3 Project, TEAM Canada. "Intellectual Framework Version 2.0," 6.

transaction. As a matter of fact, by definition, external resources are outside the scope and control of a system that does not include them, so that, due to the dynamic nature of digital information, their location or content may be unpredictably altered. Therefore, even if the presence of links to external resources does not concern whether an e-mail is entitled to be declared as a record, at the least it is necessary to adequately describe the links in an e-mail before it may be captured by a recordkeeping system.[23]

2. With regard to the connection with an action, by reason of which the e-mail has been produced or received, the main problem is that an e-mail can refer to several transactions or can result from the merging of several previous e-mails. Such situations can be kept under control by resorting to traditional diplomatic and archival analyses, possibly applying multiple classifications to the same e-mail. Multiple classifications are not inconceivable, but an extensive use of them is not a good records management practice, because the intellectual control of a recordkeeping system becomes more troublesome in the long run. Besides, if an e-mail contains information pertaining both to the activities of which the recordkeeping system is intended to give evidence and to private activities of staff persons, the very management of the e-mail is likely to give rise to serious problems, since any disclosure of its content might result in a legal action against the corporation. It is not possible to take for granted that the content of every corporate e-mail is always exclusively relevant to the business of the organization, and it is up to the recordkeeping system staff, and especially to the records manager(s), to enforce adequate policies and solutions intended to ensure that this is the case.[24]

3. The detection of an archival bond that connects an e-mail record to all the other records of a recordkeeping system always implies at some point the intervention of a human being (at least to monitor the operations carried out by means of automated systems, which of course may not be held responsible for their choices) and the actions to be performed are not qualitatively different from those done when handling paper correspondence. Nonetheless, the analysis is far trickier because the structure of an e-mail (seen as a digital object) can be quite complex: often a single e-mail may be associated with more than one record or, on the contrary, a single record may need more than one e-mail to be sent (e.g., more than one e-mail may be required to forward a single report with many large attachments).

4. The actual identification of the persons who are involved in the creation of an e-mail can prove to be impossible because an e-mail, to be delivered, needs only a sender address (not a sender that is also a physical or a juridical person) and a recipient address (not an addressee).

[23] Cf. recommendation 3.5.1, 40.

[24] When it is particularly important that e-mails only convey information relevant to one or more business processes, a corporation should write one or more ad-hoc policies that detail how to create the records intended to support these processes. If applicable, the provisions of the policies might also be enforced by setting up controlled workflows (e.g., through the use of prescribed forms or automatic content analysis tools, which are able to send warnings when something that is seemingly unsuitable is detected)

Just like paper mail, where the identification of the persons that have taken part in the compilation of a record may be sometimes troublesome or even impossible, e-mails are based on a technology that has not been designed to address this issue and, in addition, are also able to give misleading information. As a matter of fact, sender and recipient addresses may be unreliable, since it is possible to alter the relevant metadata to hide the real sender and recipient addresses[25] (even if, obviously, whoever receives an e-mail is present in the list of the addressees). The problem is so serious that it has not gone unnoticed by professional communities that usually do not deal with records management (e.g., IT specialists, lawyers). Hence, appropriate procedures must be established to identify the persons who/which have produced an e-mail purporting to be a record.

5. In relation to the five contexts taken into consideration by InterPARES researchers (i.e., juridical-administrative, provenancial, procedural, documentary, technological),[26] e-mails per se provide only some metadata relevant to the technological context. Just as in the case of paper mail, only proper targeted actions can guarantee the availability of the information required to adequately describe the five contexts.

From this short examination intended to consider e-mails in the light of the requirements the InterPARES Project has defined for records, it may be inferred that a careful preliminary analysis is needed to ensure that the e-mails that have been created in pursuance of the business of an organization, and therefore are associated with records, are correctly handled in the corporate recordkeeping system. Otherwise e-mails, instead of being "an opportunity... to do something about the management and archiving of electronic documents in general",[27] more likely will represent the kiss of death for any records management program.

2.2 TAXONOMIES

It is helpful to try to group together e-mails according to categories distinguished by characteristics that are to be taken into consideration when a recordkeeping system is being designed and implemented.

A first differentiation is determined by the overall direction of transmission: e-mails can be divided into outgoing e-mails (i.e., those that are sent by an organization to a recipient external to the organization), incoming e-mails (i.e., those coming from the outside that are received by an organization) and internal e-mails (those that are sent by an organization and

[25] As noted by Giovanni Michetti (e-mail message to the author, May 20, 2010), however, even without any alteration the address may be misleading: one might deliver a message using the account of another person (just as it may happen when paper mail is used); or one might send a message to a specific address, associated with a specific individual, and this message could actually be addressed to another individual (redirection), just as it may occur in a more traditional environment (e.g., consider the case of two different letters, i.e. two different records, addressed to different individuals using the same envelope, i.e. only one address).

[26] For an explanation of the five contexts, cf. L. Duranti and R. Preston, eds. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records* (Rome, Italy: ANAI 2008), 424-425.

[27] Cf. 6, fn17.

received by the same organization). This distinction is important in relation to the level of intellectual control that the entity that governs a recordkeeping system is able to exert. Such a level is in principle higher for outgoing and internal e-mails, because in these two cases the recordkeeping system staff can apply directions and rules intended to configure an e-mail in a given way, so that all the structures and metadata needed to appropriately manage the e-mail are present. On the contrary, the recordkeeping system staff cannot usually govern the process of production of incoming e-mails, since they are created by external entities.[28] Nevertheless, the physical or juridical person responsible for a recordkeeping system can attempt to orientate such a process of production; for example, by specifying instructions for the creators of incoming e-mails or, in some specific cases, even by preventing e-mails that do not comply with given requirements from being captured by the recordkeeping system.

Another useful division is that between corporate e-mails and non-corporate e-mails. Such a taxonomy is self-evident, trivial and, in theory, ought not to be specified: it is obvious that e-mails containing information that is unrelated to the business of an organization must not enter the recordkeeping system of that organization. Unfortunately, many people create e-mails where information concerning the transactions of the organization for which they are working are mixed with references to other situations (maybe the most classical example is that of an e-mail, the content of which pertains to the activities of the organization, but also includes information that only bears on the private sphere of the individual who has produced the e-mail and is unrelated to any aspect of the corporate business). It is paramount that all the e-mails that are declared records are exclusively by-products of the institutional activities of the entity to which the recordkeeping system refers. It is easy to articulate such a tenet, but to enforce it may be far trickier: appropriate measures must be taken to reach such a goal.

E-mails can be also categorized on the basis of their function in a thread. An e-mail sent in reply to another e-mail, or an e-mail by means of which a former message has been forwarded, is marked with a special class of headers, thanks to which it can form together with the first e-mail a "thread".[29] Of course, any e-mail that is sent in reply to, or forwards an e-mail of, a thread will be automatically added to the thread. Nevertheless, to be added to a thread, an e-mail must be produced by someone who uses the editing functions "reply" or "forward" (which every e-mail client or webmail application possesses). According to such a taxonomy we can single out primal e-mails (all the messages that have not been sent in reply and have not been forwarded; at the beginning of any thread there is always a primal e-mail), forwarded e-mails and e-mails sent in reply. This subdivision may be useful to detail specific rules affecting the process of composition of e-mails in view of their possible capture and handling in a recordkeeping system. Other practical groupings include e-mails containing only one part, or composed of more than one part, and classes of e-mails defined on the basis of the possible

[28] Cf. below 17-21. With reference to specific services or highly structured transactions, the external creation of an e-mail may be controlled by an organization by means, for instance, of forms placed in an HTML page or pre-defined templates. As a rule, only a small part of incoming e-mails may be managed in this way.
[29] G. Pontevolpe and S. Salza, 16.

values of the "Content-Type" header, while possible MIME subtypes present in the "Content-Type" header can be used to distinguish sets of e-mails that share common attributes that can be significant for records management.

Of course, many other such taxonomies could be conceived. All the taxonomies mentioned above are not grounded on a strict theoretical analysis of e-mails as digital objects. They are of a practical nature and some of them have been proposed just to facilitate the formulation of recommendations for the correct treatment of e-mails in the following section.

2.3 RELATIONSHIP BETWEEN RECORDS MANAGEMENT AND LONG-TERM PRESERVATION

In this paper, proposed recommendations for e-mail records management have been divided according to the two main phases of the records life cycle. Most of them relate to the period of time during which e-mail records are current or semi-current; namely, when the records creator uses them above all to carry out its ordinary business activities, by reason of which the records have been produced. A smaller set of recommendations refers instead to e-mail records selected for long-term preservation, when they become historical records transferred to the custody of a trusted preserver and serve no longer to further, on a regular basis, work processes, but are preserved mostly for cultural reasons and/or in view of possible sporadic occasions in which they may be again used for reference and action. Therefore, it is not without reason that the recommendations proposed for current management are more numerous than those related to long-term preservation. First and foremost, several recommendations are intended to establish good practices to create e-mail records intended to be integrated in a recordkeeping system. Second, the phase of capture is paramount to accomplish several operations aimed to enable records managers, archivists and other clerks and professionals to properly handle e-mails in a recordkeeping system and possibly afterwards in a digital archives (analysis of the intellectual and physical structure of e-mail records to correctly include them in a recordkeeping system, appropriate treatment of all the digital components of an e-mail, classification and so on).

Recommendations for long-term preservation of e-mail records focus on aspects of management that need to be modified when records become historical (e.g., choice of strategies, change of formats) or on actions that, even if they may take place when e-mail records are still current, are nonetheless totally and exclusively intrinsic to long-term preservation (e.g., use of given formats, organization of procedures for migration).

2.4 ESSENTIAL AND ADVISABLE RECOMMENDATIONS

The recommendations contained in the following section may be either *essential* or *advisable*. A recommendation is *essential* when compliance with it is presented as a necessary condition to enable organizations to appropriately capture and handle records sent by e-mail in their recordkeeping systems and then, if applicable, in their digital archives. A recommendation is *advisable* when its implementation, even if it is not deemed to be indispensable, can bring

about more effective management practices to govern e-mail records. Such a terminology is along the lines of the differentiation between "mandatory" and "desirable" requirements featured by MoReq2.[30] This paper draws on MoReq2 also in relation to the use of key words to distinguish the recommendations: an *essential* recommendation is indicated by the presence of the verb "must" in the main clause(s), while the verb "should" always appears in the main clause(s) of an *advisable* recommendation.[31] Finally, the statement of a recommendation is always in italics, while the text that explains the rationale of the recommendation is normal.[32]

3. RECOMMENDATIONS FOR E-MAIL RECORDS MANAGEMENT AND LONG-TERM PRESERVATION

3.1 INTEGRATION WITH RECORDKEEPING SYSTEMS (RKS)

The foremost guiding principle for any organization planning to set up a proper e-mail records management program is the integration of all the records conveyed by e-mail in the recordkeeping system (or in the one in the course of development, if no RKS has not been established yet) of the organization. Without such a move it is unavoidable, in concrete terms, to handle e-mails as if they were a records series, but "electronic mail, in and of itself, is not a single record series," as D.A. Wallace argued in 1998.[33] E-mail records can refer to any function, subject or activity, so that if they are kept as a de facto series, e-mail records form a separate sector in the overall RKS, to the detriment of any good records management practice. An example of a negative consequence is a series comprising records of the same form or a case file, containing all the records produced in the course of one or more given transactions, which is likely to be split, as a minimum, into two parts, each of which respectively belongs to an isolated domain (in the simplest scenario, the e-mail management system and another IT system that governs all the other digital records). Moreover, if e-mail records are gathered together in an unconnected environment, an organization is forced to operate at least two RKSs, which entails remarkable risks of duplication, fragmentation and difficult identification of the archival bond, non-compliance with the rules and policies in force and, finally, also as a result of all that has been mentioned above, the loss of the intellectual control over the records.

Therefore, the first recommendation can be now articulated:

[30] European Commission, "Model Requirements for the Management of Electronic Records (MoReq2)", Project Consult , 2008, 9. http://www.project-consult.net/Files/MoReq2_body_v1_0.pdf (last accessed March 20, 2010).
[31] European Commission, "Model Requirements for the Management of Electronic Records (MoReq2)", 9.
[32] In MoReq2 is quite the opposite: the text of the requirement is normal, while the rationale is written in italics.
[33] State of Wisconsin, Email Policy Task Force, "Literature Review- Professional standards, academic best practices", Department of Administration of the State of Wisconsin, 1. www.doa.state.wi.us/docs_view2.asp?docid=6079 (last accessed March 23, 2010).

3.1.1 *All the e-mail records made or received, and kept by an organization must be integrated in the RKS of the organization.*

The inclusion in a RKS, obviously, makes it easier to separate the e-mails pertaining to the business processes of the organization from those unrelated to them, and is conducive to increasing the ability to recognize the intellectual structure of an e-mail record, which several times does not match the physical arrangement of the electronic components that make up the e-mail understood as a digital entity.

It is not surprising, then, that much of the relevant literature explicitly recommends such an approach.[34] Boudrez points out that "export of e-mails and attachments to a folder structure outside the e-mail system is required for several reasons,"[35] while the Digitale Duurzaamheid program states that "email messages...are to be preserved on a centrally managed system and not on the computers or in the personal files of the individual users."[36] As long ago as 1998 the *Govern Recordkeeping Manual* of the New South Wales Government prescribed that "Electronic messages required as evidence of substantive business activity should be captured directly into an electronic recordkeeping system,"[37] and, in 2007, Queensland State Archives confirmed that "Management of emails that are public records should not occur in isolation from the management of paper-based or electronic records."[38] Huth wrote in 2002 that:

> Records needed to support program functions should be retained, managed, and accessible in existing filing system **outside the e-mail system**[39] in accordance with the appropriate program unit's standard practices.[40]

Of course, the records management standards MoReq2 and DoD also provide that e-mails be captured by ERMSs.[41] There are plenty of sources that advise integration. In a sense, it is obvious that e-mail records cannot be kept outside RKSs, since the most important incoming

---

[34] The organization and streamlining of appropriate workflows is a powerful system to control any kind of record: the use of workflow management tools and strategies may be an important element of a corporate plan aimed at integrating email records in a RKS. With regard to workflows, cf. below 19-22.
[35] F. Boudrez, *Filing and Archiving E-Mail*, 8.
[36] Digital Preservation Testbed. *From digital volatility to digital permanence. Preserving email* (The Hague: Digital Preservation Testbed 2003), 68.
[37] State Records Authority of the New South Wales, "Government Recordkeeping Manual. Rules. Polices. Policy on Electronic Messages as Records", 1998, #3.1. http://www.records.nsw.gov.au/recordkeeping/government- recordkeeping-manual/rules/policies/policy-on-electronic-messages-as-records (last accessed March 23, 2010).
[38] Queensland State Archives, "Managing Emails that are Public Records", 2007, 3. http://www.archives.qld.gov.au/downloads/emails_that_are_public_records_policy_and_guideline.pdf (last accessed March 23, 2010)
[39] Bold characters in the source.
[40] G. Huth, *Managing E-mail Effectively*, Archives Technical Information Series 62 (Albany, NY: New York State Archives 2002), 38. http://www.archives.nysed.gov/a/records/mr_pub62.pdf (last accessed March 23, 2010).
[41] European Commission, "Model Requirements for the Management of Electronic Records (MoReq2)", 79-84. Department of Defense, *DoD 5015.2 STD, Design Criteria Standard for Electronic Records Management Software Applications* (Assistant Secretary of Defense for Networks and Information Integration / Department of Defense Chief Information Officer 2007), 40-41. http://jitc.fhu.disa.mil/recmgt/p50152stdapr07.pdf (last accessed March 23, 2010).

and outgoing flows of digital records take place by e-mail technology, so that an RKS that does not include e-mail records is an RKS without most of the corporate records.

Nevertheless, to manage e-mail flows, a great deal of small- and even middle-sized organizations normally make use of one of the numerous e-mail client and webmail applications available on the market.[42] There is no point in denying that to use an integrated system, the architecture of which has been designed with the goal to ensure that e-mails can be routed – automatically or on request – to and from an RKS for possible capture and/or dispatch, is the best option for records management purposes, but it is unrealistic to demand that all the organizations, even the smallest ones, set up such systems. In this respect, it is worth considering the arrangement of plug-ins for the export of e-mails and their attachments from a client program to an RKS and vice versa. Some research projects, such as DAVID,[43] Digitale Duurzaamheid[44] and Paradigm,[45] have tested the viability of such a solution. This strategy tries to conciliate the implementation of more stringent records management procedures concerning e-mails with existing practices and available financial, technological and human resources, so as to minimize as much as possible the resulting impact of the changes in work environments: users' point of view is always to be taken into consideration, because, as Wallace has pointed out, the end-user is the one who "manages electronic mail."[46]

The role of end-users is also paramount in relation to another main issue in e-mail management: how to restrain personal accounts from swallowing e-mails that belong to an organization. It is evident that this problem directly affects the integration with RKSs. Usually, an organization uses many accounts. They can be very numerous in large entities that receive millions of e-mails each day.[47] As previously described,[48] there are individual business accounts and business accounts related to functions or offices. Broadly speaking, for the purpose of furthering the inclusion of e-mail records in RKSs, the fewer corporate e-mails that pass through individual business accounts, the better. We can therefore articulate a second recommendation:

3.1.2 *In an organization, business activities must be carried out by using, as much as possible, corporate accounts that have been associated with particular offices or have been established to fulfill specific functions, and can be freely accessed at least by given members of the RKS staff.*

In this way it is easier to direct to RKSs e-mails that are associated with records. Of course, the most radical solution of the problem is the elimination of individual business accounts, but such an organizational schema is hardly ever feasible, because the role that

[42] G. Pontevolpe and S. Salza, 8.
[43] F. Boudrez, *Filing and Archiving E-Mail*, 8-11.
[44] Digital Preservation Testbed, *Technical Description TestbedXMaiL* (The Hague: Digital Preservation Testbed 2005), 6.
[45] Paradigm, "Workbook on Digital Private Papers. Exporting email from Microsoft Outlook email clients". http://www.paradigm.ac.uk/workbook/accessioning/microsoft-outlook.html (last accessed March 23, 2010).
[46] State of Wisconsin, Email Policy Task Force, 1.
[47] The Sedona Conference, *The Sedona Canada Principles. Addressing Electronic Discovery* (The Sedona Conference 2008), 2.
[48] Cf. 3-4.

individual business accounts play in corporate activities is usually too significant. Nevertheless, to comply with such a recommendation not only reduces possible losses of records, but also avoids the chance that given categories of records may be seen by persons who are not entitled to do so. A third recommendation is strictly connected with the former one:

3.1.3 *An organization must draw up a list of the e-mail accounts enabled to directly interact with the RKS.*

In other words, the staff of an organization must possess full awareness of the connections between the RKS and the whole set of corporate e-mail accounts. An ideal situation could seem to be one where there is a single entry point to the RKS through one e-mail account; however, apart from the improbability of such an arrangement, as long as it has the control over all the gates to the RKS, an organization can easily succeed in monitoring incoming and outgoing traffic by means of appropriate applications.

To perform to the largest extent work activities through business accounts related to functions or offices, it is indispensable to dictate relevant rules and resort to Business Process Reengineering (BPR).[49] These goals can be reached by means of an array of instruments, which include policies, modeling of transactions and designs of document templates. Therefore, a fourth recommendation can be now enunciated:

3.1.4 *An organization must select tools and strategies intended to facilitate the integration of e-mails records in the RKS.*

This is proposed as an essential recommendation since some simple measures, such as to draw up a policy, are within the reach of any organization. On the contrary, other approaches, such as BPR, require at least a certain amount of resources and therefore cannot be implemented by all organizations.

In the event that a physical or juridical person has no formal RKS or, for whatever reason, is unable to integrate the e-mail client or webmail application with its RKS, the only viable solution is to replicate in the e-mail environment the way in which all the other digital and analog records produced by that person are organized (i.e., it is necessary to apply the same classification scheme both to e-mail records and to all the other records; the same files are to be created in all the records management environments used by the person; etc.). At least from an intellectual point of view, the integration must be pursued. In principle, the situation is not different from managing a whole of records that are in part digital and in part analog (e.g., paper records). Of course, this increases dangers of misalignment and loss of control, so careful monitoring of the records becomes even more essential to prevent disarray. However, if a formal RKS exists, a solely intellectual integration should be only a transitional solution: such a scenario is highly likely to be untenable in a long-run perspective. Of course, only records

[49] BPR is a discipline that deals with the analysis, design and organization of work processes for the purpose of optimizing them.

creators whose organizational structure is very simple and produces just tiny amounts of records can work without a formal RKS.[50]

3.2 CAPTURE AND FILING

Capture is an action by means of which one is able "to save a particular instantiation or state of a digital component or group of components,"[51] as defined by the InterPARES 2 *Glossary*. This meaning can certainly be applied to the operation of inclusion of e-mails in an RKS because e-mails are objects that are physically made up of digital components. The final goal of such an inclusion is to set aside records that have been sent by e-mail to an RKS, so that all the intellectual relationships between each record and all the other records kept in the RKS and between each record and the whole of the records kept in the RKS are clearly expressed or, in other words, so that the archival bond of each record can be made explicit and understood. Therefore, the step of inclusion is both physical and intellectual, and the presence of an archival bond always implies that a record will hold an unique position in the RKS. We can term the intellectual aspect of the process of inclusion "filing" because according to the definition provided by InterPARES 2,[52] "to file" means:

> To set aside a made or received document among the records that participate in the same action/affair or relate to the same person or subject, so that they may be retrieved for action or reference.

Furthermore, the manifestation of the archival bond of a record at all times clarifies to which action or subject the record relates, although there may be sometimes a practical coincidence (but never an intellectual one) between a record and a file (in case a transaction should require just one record to be carried out, or only one record should be created in relation to a given matter).

The action of "capture and filing" an e-mail record must be carefully planned, since potential mistakes could be irreparable. The first recommendation of this section is a direct corollary of recommendation 3.1.3:

3.2.1 *The operation of capture and filing must be performed only through the e-mail accounts enabled to directly communicate with the RKS.*

This recommendation is aimed to avoid the unrestrained proliferation of inward and outward flows that may complicate an already difficult process. Moreover, it serves to prevent individual business accounts from randomly becoming gates to an RKS. In cases where an e-mail record has been received in an account not enabled to directly communicate with the RKS,

[50] In any case, compliance with at least the essential requirements of these guidelines is deemed to be indispensable even for such records creators, if they wish to properly and effectively manage their e-mail records.
[51] L. Duranti and R. Preston, eds., 777.
[52] L. Duranti and R. Preston, eds., 789.

the record is to be handled in compliance with specific corporate rules (e.g., a policy might state that the e-mail is to be forwarded to an enabled corporate account).

E-mails can be made up of many digital components. In an e-mail record, the physical organization of those components might not mirror the intellectual structure of the record. For example, consider a letter with some enclosures: in an e-mail, in theory, both the letter and its enclosures can be attachments to the part of plain text that is displayed online together with the headers and which can contain just some simple words (e.g., "here is the letter!") or even be empty.[53] Of course, all organizations should define as precisely as possible the intellectual and physical structures of the e-mail records that will be included in their RKSs and try to enforce as much as possible compliance with it, but it is evident that, especially in relation to incoming e-mails, organizations cannot control the design under any circumstance. These observations are the basis of the next recommendation:

3.2.2 *The operation of capture and filing must include e-mail records in the RKS so as to always highlight their intellectual structure. The information about the physical order of the digital components that make up an e-mail record must be always preserved as metadata in the record profile of the e-mail record.*

This recommendation might appear to be unclear and vague, but what it implies is to provide and preserve adequate information about the intellectual and physical structure of a record. When dealing with paper records, similar cases are far less frequent, but may sometimes occur (e.g., when all the other parts of a record are physical attachments of a routing slip or an empty cover letter, just containing the data of the sender and addressee: if the intellectual structure of the record is considered, the routing slip or empty cover letter are to be regarded as attachments, since they report only information intended to support the transmission of the record). With regard to both e-mail and paper records, the intellectual structure of a record may be highlighted through the use of appropriate metadata elements (e.g., indication of the part of the record in relation to which all the other parts are to be considered enclosures, subject, type of record, list and description of the attachments) and possibly of specific methods of visualization (for instance, if an RKS displays the classical tree visualization, there might be a solution to let a user easily understand the intellectual structure of an e-mail record by this kind of representation).

The process of capture and filing can be really troublesome, owing to the huge range of conditions and configurations that may characterize especially incoming e-mails. Records

[53] The headers and the part of plain text displayed just below them are the central section with regard to the physical organization of the e-mail both because they contain many metadata elements essential for the management of the e-mail during sending, transmission and reception, and because they are the central hub for all the links to all the other digital components of the e-mail. On the contrary, according to this example, the letter, even if it is an attachment with relation to the physical organization, actually reports all the information necessary to understand the structure of the record and to identify which parts are to be considered enclosures, as well as including many elements intended to describe the intellectual functions and characteristics of the record (e.g., the persons involved in the creation of the record, the reasons that have led to the creation of the record, the connections with other records kept by the corporation, etc.).

managers and archivists, in cooperation with all the other professionals and members of the RKS staff, must strive to restrict the variety of situations with which they have to cope. First of all, they have to decide which metadata elements should be associated with every e-mail record and in which format they should be. The verification of the presence of such elements in an e-mail must be coordinated with the action intended to identify the archival bond of the e-mail. The level of intellectual analysis required by these activities apparently needs a certain amount of time to be carried out. If we think of the enormous number of e-mails that must be examined for a possible capture in an RKS on a daily basis, it is easy to infer that the resulting workload, if not properly managed, can completely disrupt the most efficient organization. Therefore, it is imperative that organizations automate the process of capture and filing (as well as any other phase of the handing of e-mail records, even if capture and filing probably bring about the most perceivable and immediate difficulties) as much as possible. Records managers must be capable of availing themselves of the most advanced techniques of content analysis and automated management for the purpose of expediting the procedure of capture and filing, since it is unrealistic to assume that organizations are willing to allocate more than a limited share of time and human resources to records management. However, "as much as possible" at the same time does not mean "more than is necessary". The process of capture and filing cannot be completely automated. Machines cannot substitute records managers and other information professionals, primarily because of their intrinsic inadequacy to satisfactorily solve a gamut of complex problems: there is no software or machine-driven evaluation that can overcome the ambiguities of interpretation and meaning that make it difficult for a human being to analyze a document and classify a record. Moreover, obviously no machine can be held accountable for its work. On the basis of these reflections it is therefore possible to articulate two other recommendations:

3.2.3 *The process of capture and filing of e-mail records in an RKS should be automated as much as possible.*

3.2.4 *In cases where the process of capture and filing is partly automated, the RKS staff must implement procedures that let humans audit the work carried out by machines and guarantee the accountability for the decisions that have been made.*

The first recommendation is advisable, since there are small organizations that might not need any automated procedure. However, in practice it is unthinkable that large organizations may do without a machine-aided approach and, in this respect, records managers and archivists must not be only passive onlookers, restricting themselves to giving some advice to IT professionals about the characteristics an RKS software must possess; they must behave in a proactive fashion, using their expertise and making the most of traditional tools that have been supporting their activities for decades. Records managers are by now used to modelling workflow processes and therefore are arguably the professionals best positioned to govern the automation of recordkeeping systems and to give advice to all the specialists involved in the development of automated procedures for records management. Moreover, many archivists have

at least some knowledge of diplomatics, the discipline that deals with the study of the formal aspects of records. These skills can turn out to be amazingly useful to facilitate human and machine work. In fact, as already said, digital documents, and in particular e-mails, are difficult to be appropriately analyzed because of both their number and their kaleidoscopic and extremely irregular array of forms and configurations. In that sense, only an alliance between human intelligence and computing power can master them because, on the one hand, machines can help human beings to manage a gigantic quantity of e-mails – for instance by means of metadata harvesting, automatic parsing and proposals of classification – on the other hand, human beings must aid machines not only by validating final results, but also by reducing the complexity and irregularity of the objects that must be evaluated. By resorting to diplomatics, archivists and records manager can design regular documentary forms intended to characterize e-mail records (as well as other kind of records, if appropriate) and ensure, in cooperation with IT professionals, that they are actually applied to, the largest possible extent, to the e-mails that are included in RKSs. This way diplomatics can still play an important role in our society, not by letting scholars recognize the juridical nature of archival documents, but by inserting in documents planned to be declared records recurring formal elements that are able to be easily identified by both machines and human beings. In other words, diplomatics could rejuvenate and become a discipline no longer mainly aimed to strengthen the evidentiary value of records as it was the case in medieval and modern European chanceries, but to expedite records management operations.

In concrete terms, how can all this be enacted? Technology is the least of the concerns: there are plenty of ways to create records that feature regular structures and fixed forms. To this purpose, XML Schemas are outstanding because they are both machine- and human-readable and are based on a semantics that is public and independent of any specific application. The design of e-mail templates in XML format for records management and long-term preservation has also been proposed by the researchers of the project Digitale Duurzaamheid.[54] Nevertheless, the real troubles are others: first of all, an organization is supposed to have the control of the making of outgoing and internal e-mails, but it is not always possible to enforce the adoption of e-mail templates for incoming e-mails, which are normally a substantial flow. Secondly, how many e-mail templates does an organization require? There are thousands of different transactions and situations and one can never fully anticipate what kinds of contents and documentary forms will be needed when creating records. To produce and handle thousands of templates, many of which would become quickly outdated, is a cumbersome task and would compound the overall workload of the personnel of an organization. In the first place, it is not always necessary to prepare complete e-mail templates. Just the introduction of some regular patterns in the documentary form of records may be of great use to machines and RKS staff. Then, an action intended to normalize the documentary form of records must be necessarily coupled with the reengineering of workflows and business processes. As in the case of

[54] Digital Preservation Testbed. *From digital volatility to digital permanence. Preserving email*, 61.

diplomatics, this is not a new skill for information professionals: records managers especially are particularly adept at studying workflows to understand how records are created in a given environment, preparing management tools (i.e., schemes of classification and retention and disposition schedules) and devising opportune modifications. Through modelling and workflow analysis, records managers can discover which are the most widespread typologies of records and, as appropriate, suggest convenient modifications to their configuration. If need be, such an activity can be focused on e-mail management for the purpose of building templates suitable for e-mail records created as by-products of given transactions. Moreover, a thorough and attentive examination of business processes can yield other positive outcomes, such as a description of the metadata elements (or, if applicable, even of metadata values) that are to be present in specific kinds of e-mail records or a list of the attachments that must be appended to a record. Finally, BPR can also be an effective method to direct as much as possible the creation and configuration of incoming e-mails. For instance, a workflow could be planned so that incoming e-mails mandatorily contain definite values of metadata (e.g., a specific address; the subject) or comply with a particular arrangement of the digital features (e.g., an enumeration of the attachments to be appended to an e-mail; file formats accepted for the attachments of an e-mail). In some specific business processes a highly structured workflow could require external users to fill out a form in the corporate website (which is to be subsequently captured by the RKS) instead of sending an e-mail through their client or webmail applications; records managers might arrange pre-established templates intended to support the organization of automated workflows where e-mails go through a sequence of routine steps. Rule enforcement is easier if it concerns a single work process, because one can issue directions in a more flexible manner and at the same time focus them on more definite targets. However, when designing business processes, exhaustiveness is not necessary and to try to obtain it at any cost is a mistake. The overall goal is to better govern records flows and, within such a framework, BPR proves to be successful if records managers have succeeded in identifying, describing and, if need be, modifying all the work processes that feature distinguishable patterns and bring about the creation of a substantial share of records. Of course, the effort must be commensurate to both expected advantages and available resources and supported by the highest hierarchical level of an organization.

A strategy where advanced automation, business process reengineering, management-oriented diplomatics and human supervision interact would enable the implementation of the "combination of manual and automated classification methods,"[55] which according to James Santangelo, is the best option to cope with gigantic quantities and unstructured data and forms. To pursue this approach can be problematic, not so much because of the lack of suitable technological instruments, but because of the difficulties to establish an effective cooperation among different professionals and corporate roles in an organization and gain the needed

[55] J. Santangelo, "Rise of the Machines: The Role of Text Analytics in Record Classification and Disposition", in *Information Management Journal* 43, no. 6 (November-December 2009): 26. http://content.arma.org/IMM/Libraries/Nov-Dec_2009_PDFs/IMM_1109_rise_of_the_machines.sflb.ashx (last accessed March 26, 2010)

consensus of executive levels, workers and other stakeholders to start projects of this sort. Therefore, as a result of such considerations, two more recommendations can be formulated:

3.2.5 *Templates and recurrent patterns intended to design regular documentary forms to be applied to e-mail records should be arranged by the RKS staff to expedite automated and human intellectual analysis of e-mails with a view to their capture and subsequent handling in the RKS.*

3.2.6 *The RKS staff should resort to Business Process Reengineering modeling and analysis to understand which kinds of records are needed to carry out given transactions and which characteristics and sets of metadata elements and values must be present in such records.*

These recommendations are only advisable since such approaches are primarily meant to deal with considerable e-mail flows and reduce the time and resources required to appropriately handle them. A small entity could also choose to avoid the initial effort to implement them, although they may trigger positive effects in any work environment. However, they are actually essential for a large organization.

Regardless of any method to deal with e-mail flows for the purpose of performing capture and filing, it is up to each organization to exactly define which intellectual and digital characteristics are needed to make up e-mail records and which metadata elements (and, if applicable, which metadata values) have to be contained in such records. At the same time, some features and pieces of information should be present in any e-mail intended to be a record. The MoReq2 standard requires the extraction of a series of metadata (for instance "e-mail date sent", "recipient(s)", "any copy recipient(s)", "subject line", "sender") "to the extent that these are present."[56] Nevertheless, these are all metadata related to an e-mail as a digital object and, as such, are by no means sufficient to properly describe an e-mail record, also because they have not been designed for that purpose. As one can infer from section 2.1,[57] the information usually found in e-mail headers is often unsuitable for manifesting the characteristics of a digital record that have been identified by InterPARES 1 and 2. Therefore, adequate actions are required to ensure that when e-mails are being captured in an RKS, all the metadata and features essential for the establishment of the archival bond are available. Failure to find them in an e-mail as a rule must lead the RKS personnel to accept the e-mail record in the RKS regardless since, on the whole, the rejection of an e-mail record is not a viable option (especially with regard to incoming e-mail records). As a matter of fact, not capturing a record is an action that can be accomplished rarely, since the RKS staff do not have the authority to dismiss any record created as a by-product of a transaction. This option is possible only if a policy approved by the highest hierarchical level of an organization provides that a record that does not possess given

[56] European Commission, "Model Requirements for the Management of Electronic Records (MoReq2)," requirement 6.3.5.
[57] Cf. 8-10.

characteristics is not to be included in the RKS. Normally, the RKS staff must only annotate the missing metadata and/or features in the profile of an e-mail record.

In consideration of what has been said above, the following recommendations can be now articulated:

3.2.7 *The RKS staff must define the list(s) of digital and intellectual features needed to make up e-mail records and the list(s) of metadata that must be associated with an e-mail record.*[58]

If need be, more lists must be drawn up and the lists can be differentiated according to the kind of record and of transaction to which the record pertains.

3.2.8 *Any record sent by e-mail must be included in the RKS as a separate entity.*

For the purpose of developing appropriate management strategies to handle e-mail records and the RKS where they are held, each record must exist as a separate entity in the RKS. Of course, this recommendation also implies that the RKS, in any event, cannot accept file formats that merge more than one e-mail record into a single digital entity. A single e-mail that is composed, for instance, of two different records cannot be included in the RKS just the way it is: the RKS staff must separately include each of the two records in the RKS (in such cases there are one or more parts of the e-mail – as a minimum, the headers – shared among all the records associated with the e-mail: it is therefore essential that these parts are appropriately integrated with each record, according to the characteristics peculiar to every single shared part).[59] Although there are no theoretical problems with regard to the breakdown of an e-mail into multiple records, such an operation inevitably requires and consumes human and technological resources. Therefore, the following advisable recommendation can be articulated:

---

[58] A list of digital features includes the file formats that are to, or may be used to, create an e-mail as well as the kinds of technological objects that may be part of an e-mail, such as digital signatures and other encryptions, specific data types, links, html code, stylesheets, etc. A list of intellectual features enumerates which elements must or may make up the intellectual structure of an e-mail record (e.g., specific templates and/or fields, predefined headings, special kinds of attachments that are to be found in conjunction with given categories of records, particular stylistic rules, etc.).

[59] Also, MoReq2 envisages that a single e-mail may be associated with more records, and that in this case it must be possible to separately capture each record. Cf. European Commission, "Model Requirements for the Management of Electronic Records (MoReq2)":

- requirement 6.3.8:
  Where an e-mail and its attachment(s) are captured at the same time but as separate records, the resultant records should be linked automatically by the ERMS.
- requirement 6.3.7:
  The ERMS must allow a user to choose how to capture an e-mail message with attachment(s) as:
  - the e-mail message only, without attachment(s);
  - the e-mail with its attachment(s), as one record made of linked components;
  - the attachment(s) only, each or any as individual records.
- requirement 6.3.9
  Whenever an attachment is captured as a separate record, the ERMS must require appropriate record metadata values to be captured and/or entered for it.

3.2.9 *Any e-mail should be associated with one and only one record.*

On the basis of this recommendation, every e-mail should correlate with only one record, in the sense that all the digital parts forming the e-mail (e.g., headers, plain text parts of the body, attachments) should pertain to the same record. Of course this does not mean that one should not keep track of the physical connections of the parts of an e-mail, the content and metadata of which refer to more records, or even worse to split a single part. What is advocated here is a preventive action (the implementation of which should, however, be due for outgoing and internal e-mails and should be attempted, insofar as it is possible, for incoming e-mails). In the situation in which an e-mail should actually be associated with two or more records, the only viable solution is to intellectually break down the e-mail in more records and link with any record the metadata elements and, as appropriate, digital components of the e-mail that are common to all the records.

To try to actually identify at least some of the five persons who, according to the findings of InterPARES 1,[60] are involved in the creation of a digital record, it is necessary to determine the identity of the sender and of the addressee. With regard to senders, indications of e-mail accounts alone are inadequate. The RKS staff must be able to link any account with a definite physical or juridical person. If they do not succeed in doing so, the e-mail record is to be regarded as created by an unknown entity. Hopefully, problems of identification of senders should affect only incoming e-mails. Of course, a record may be included in any RKS even if its author is unknown: for example, law enforcement agencies file on a regular basis anonymous letters and memos. However, these are abnormal circumstances with respect to both analog and digital records and the RKS staff must strive to always assign an e-mail record to a person or an organization who or which is presumed to be the sender.

To facilitate the identification, lists (e.g., databases, spreadsheets) allowing the RKS staff to link each sender account with its associated entity must be arranged. When dealing with an e-mail record coming from a new sender address, the RKS staff must try to connect the sender address with an individual or an organization who/which has actually sent the e-mail and consequently update the list(s) in conjunction with the inclusion of the e-mail record in the RKS.

3.2.10 *The sender address of an e-mail record must be associated with an identified physical or juridical person in conjunction with the inclusion of the e-mail record to the RKS. If the association proves to be impossible, the author of the e-mail record must be regarded and registered as an unknown entity.*

As explained above,[61] the RKS staff are not entitled to reject any record, if appropriate provisions do not enable them to do so.

[60] Cf. 8.
[61] Cf. 22-23.

3.2.11 *The RKS staff must ensure that one or more lists (e.g., databases, spreadsheets) to link sender addresses with the physical or juridical persons to which they relate are kept and updated. The RKS must be connected with the list(s) to make the identification of a sender easier. Procedures and means employed to check the identity of the physical or juridical person associated with a given sender address must be reported in the list(s).*

It is up to the organization (clearly, by abiding by the overall framework of legislation and all the other relevant regulations) to establish procedures to ascertain the identity of a sender and detail the means that must be employed to attain a higher level of reliability, but the RKS staff must keep track in the list(s) of such procedures and means.

As for the e-mail accounts of addressees, the main difference, in comparison to sender addresses, is the possible presence of lists of distribution in substitution of the indication of single e-mail accounts. It is a good practice to associate all the accounts of addressees with a physical or juridical person.

3.2.12 *All the e-mail accounts of the addressees of an e-mail record must be associated with an identified person in conjunction with the inclusion of the e-mail record in the RKS. If the association proves to be impossible, the addressee to which that account relates must be regarded and registered as an unknown entity.*

3.2.13 *The RKS staff must ensure that one or more lists (e.g., databases, spreadsheets) to link the e-mail accounts of addressees with the physical or juridical persons to which they relate are kept and updated. The RKS must be connected with the list(s) to make the identification of an addressee easier. Procedures and means employed to check the identity of the physical or juridical person associated with a given email account of an addressee must be reported in the list(s).*

3.2.14 *If a distribution list is in the 'to' header of an e-mail associated with one or more records, the RKS staff must strive to gather all the available documentation about the distribution list. The documentation concerning the distribution list must be kept in the RKS and must be reported in or linked with the record profiles to which it pertains.*

3.2.15 *In relation to outgoing and internal e-mails, an organization must set rules about the use of the headers 'carbon copy' and 'blind carbon copy' and corporate policies must establish when an e-mail account of an addressee can be placed in the 'carbon copy' and in the 'blind carbon copy' headers.*

3.2.16 *Recommendations 3.2.12, 3.2.13 and 3.2.14 must be applied to all the e-mail accounts and distribution lists placed in the 'carbon copy' and 'blind carbon copy' headers.*

The purpose of recommendations 3.2.14, 3.2.15 and 3.2.16 is to avoid anarchy regarding the use of distribution lists and the 'carbon copy' and 'blind carbon copy' headers. Distribution lists and blind carbon copies hide meaningful pieces of information and consequently prevent RKSs from extracting important metadata from e-mails, so that an organization must strive to make such information available in a different way. The documentation concerning a

distribution list mentioned in recommendation 3.2.14 should clarify who or which entity has compiled the distribution list, which are the e-mail accounts belonging to the list and, if possible, which physical or juridical person is associated with each account of the list. With respect to blind carbon copies (and sometimes also to distribution lists) some information is concealed on purpose, but even in these cases an organization must always specify in which circumstances this sort of copy may be used, where the information can be found and, in case the information should not be open to all the RKS staff, which corporate roles are entitled to access it. Policies and other appropriate documentation must regulate how all these tools must be employed.

Of course, recommendations 3.2.12, 3.2.14 and 3.2.16 can be rather easily enforced as it concerns outgoing and internal e-mail records, since an organization should hopefully be able to control its methods of production of e-mail records. Nevertheless, an organization may receive an e-mail that has been addressed to a distribution list and not to an e-mail account of the organization, or an e-mail where a corporate e-mail account is just one of the addressees of the e-mail or only appears in a carbon copy or in a blind carbon copy header. In these cases problems are not qualitatively different from those involved by the identification of a sender address. Ideally, all the accounts of a distribution list and all the accounts which do not belong to the organization, but are contained in addressee, carbon copy or blind carbon copy headers, should be equally associated with an identified entity (recommendation 3.2.12) and one or more lists should be set up to facilitate the connection between an e-mail account and the entity to which it pertains (recommendation 3.2.13). Again, ideally the RKS should also always report the header and, if applicable, the distribution list where a given account has been found. Nevertheless, especially for distribution lists, the association of all the accounts with identified entities rather often proves to be a task that is downright impossible or simply too demanding owing to the workload it involves. In this case, the RKS staff, apart from possible exceptions established by laws and policies, has to accept the e-mail record, but also must report that one or more e-mail accounts remain unidentified (recommendation 3.2.12).

The subject header of an e-mail can greatly help the RKS and the RKS staff to determine especially the procedural context and the content of a record and therefore can provide essential metadata elements. Moreover, it is indispensable for records management purposes that the subject of a record is properly indicated, especially for some operations such as classification and filing. Unfortunately, this task is particularly time-consuming and can make the workload far heavier for the RKS staff in the immediate future, despite bringing about many positive effects. However, it needs to be noted that in some countries (Italy, for example), where legislation provides that a registry system monitors the inclusion of records to electronic or analog RKSs belonging to public bodies, it is mandatory to indicate the subject of a record. In this respect, a meaningful subject header can support the determination of the subject of an e-mail record. Nevertheless, a subject header can be also empty or filled in with trivial words (e.g., “request of information”, “call for help”, “greetings” and so on), so that appropriate actions are required to ensure that subject headers may be useful. First of all, compliance is required with the advisable recommendation 3.2.9: as a rule, it is impossible to suitably use the subject header

if a single e-mail has two attachments, each of which contains a different record. Then, the subject header must adequately convey the transaction or the subject to which the record relates. To decrease the unavoidable workload this operation entails, BPR and diplomatics techniques, as stated above,[62] can be particularly useful to the RKS staff. As a matter of fact, through a combination of these methods it is possible to design specific e-mail templates[63] intended to be used in given transactions and featuring a predefined content for the subject header. In addition, a database (to be regularly updated), or any adequate tool containing a series of subjects to be linked with one or more transactions or activities, can be set up and connected with the RKS, to normalize subjects and support the RKS staff and any automated tool when they try to describe the subject of records sent or received by e-mail (In Italy a project to normalize the description of records subjects, named "Aurora"[64] and directed by the State University of Padua, led in 2009 to the publication of a series of recommendations. The project was also aimed at normalizing the wording of senders and addressees). An organization, by exploiting such resources, can prepare outgoing and internal e-mail records so that each of them contains a subject header that is able to effectively express the subject of the record. As for incoming e-mail records, subject headers must be not modified: in case of inadequacy of the wording of the subject line, the RKS staff has to add in the record profile a more detailed definition of the subject of the record. However, again by means of BPR and strategies based on diplomatics, given workflows, related to highly structured transactions that are particularly important for an organization, can be arranged in such a way that an incoming e-mail record is automatically sent back to the sender unless the subject header of the e-mail contains a predefined wording (to design and implement such a kind of workflow, the RKS staff obviously needs at least the consensus of the highest hierarchical level of the organization).

We can then state the following recommendations:

3.2.17 *The RKS staff must ensure that the subject of an e-mail record to be included in the RKS is reported in the record profile and that the subject meaningfully expresses the transaction or the matter to which the record relates.*

3.2.18 *The RKS staff should ensure that the subject header of any outgoing and internal e-mail meaningfully expresses the subject of the e-mail record.*

3.2.19 *The RKS staff must ensure that the subject header of any incoming e-mail to be included in the RKS is captured the way it is and added to the record profile of each record associated with the e-mail. If the subject header does not meaningfully express the*

[62] Cf. 19-21.
[63] For some examples of corporate e-mail templates, cf. State Records Authority of New South Wales, "Government Recordkeeping Manual. Guidance. Guidelines. Guideline 13 - Create and capture: Guidelines on better recordkeeping," http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping- manual/guidance/guidelines/guideline-13 , 2001, #4.2 and #4.3 (last accessed May 4, 2011), and Texas Digital Library, "Vireo Email Templates," http://wikis.tdl.org/tdl/Vireo_Email_Templates , 2010 (last accessed May 4, 2011).
[64] The State University of Padua. "Cos'è Aurora" (What is Aurora?). http://www.unipd.it/archivio/progetti/aurora/cosa.htm (last accessed March 27, 2010).

subject of the e-mail record or expresses it wrongly, the RKS staff must anyway ensure that the record profile reports an adequate definition of the subject of the record.

3.2.20 *The RKS staff should set up appropriate tools (e-mail templates, lists of normalized subjects, controlled workflows) to expedite the definition of an appropriate subject for any e-mail record and its insertion in the associated record profile and, if possible, in the subject header of the e-mail.*

Recommendations 3.2.18 and 3.2.20 are only advisable because the one-to-one correspondence between records and e-mails (cf. 3.2.9) has not been proposed as an essential recommendation, since it could be impossible to comply with it in some work environments.

3.2.21 *When an e-mail record is captured by the RKS, the RKS must ensure that date sent and date received and, if applicable, time sent and time received are added as metadata to the associated record profile.*

Date sent and time sent can be found in the date header of the e-mail. Date received and time received must be added by the RKS and relate to the date and time of capture in the RKS. Organizations can set the RKS to register other kinds of date and time: for instance, the RKS could be set to register the date received and time received related to the moment of receipt of the e-mail by a corporate account enabled to directly communicate with the RKS (cf. recommendation 3.1.3), before the actual capture of the e-mail record in the RKS.

3.2.22 The *RKS must add to the record profile of an e-mail record all the headers of the e-mail associated with that e-mail record.*

The headers contain many important metadata elements, such as the IP address of the sender, the identifier of the message and the path the e-mail has followed to reach the addressee.

As already stated,[65] the one-to-one correspondence between e-mail records and e-mails (cf. 3.2.9) is strongly advisable.

3.2.23 *The RKS staff must draw up a list of the file formats that can normally be kept in the RKS. The list must be updated on a regular basis. If an e-mail record includes an attachment in a file format that is not present in the list, the RKS staff must strive to convert the attachment into one of the file formats of the list. The list must differentiate the acceptable file formats according to the requirements established as a result of specific retention and disposition schedules and, if applicable, of other relevant provisions. On-line cross-references from the list to the retention and disposition schedules and any other relevant provision must be available. The RKS must be connected with the list, both to enable the RKS to prompt possible decisions and to make the information available for reference.*

[65] Cf. 23-24; 26-27.

This recommendation is intended to further the long term-preservation of an e-mail record or at least its usability and reliability throughout the retention period established by the relevant retention and disposition schedule. If an e-mail record includes an attachment encoded in a format unsuitable for the retention period of the record or an attachment that cannot correctly be opened and handled by the RKS and, in addition, the format at issue cannot be even converted, the RKS staff must annotate that in the record profile, unless a law or a relevant policy approved by the highest hierarchical level of the organization provides that a record containing such an attachment must be rejected or that another predefined action must be started. The list must be updated on a regular basis by an RKS staff member entitled to do so.

As said in section 3.1, the key move for the appropriate management and long-term preservation of e-mail records is their complete integration in the RKS. We may therefore enunciate the following essential recommendation:

3.2.24 *During the phase of capture and filing, any e-mail record must be assigned a unique identifier and must be classified, filed, associated with a retention and disposition schedule and handled just as any other record captured in the RKS.*

In the event that there is no formal RKS or, for whatever reason, the e-mail client or webmail application cannot be integrated with the RKS, capture and filing procedures necessarily take place in the application environment that a physical or juridical person uses to manage its e-mails. However, such procedures must be always in keeping with those performed to capture and file all the other records created or received by the person. As has already been said at the end of section 3.1,[66] the intellectual integration must be anyway pursued. This means, for example, that if the person uses a classification scheme to organize its records such a scheme has to be also applied to the records associated with e-mails; if the person assigns to each record a unique identifier, an ID must be likewise given to each e-mail record and the ID value must be determined by taking into account all the records produced and set aside by the person. Many requirements specified in this section and in the following ones can be applied even in the absence of any formalized RKS, although in this case their implementation is evidently to be carried out in a more elementary way. Requirement 3.2.2 still holds even if no formalized RKS is available, because the intellectual and physical structure of an e-mail record must be always correctly described and represented. However, if the physical order of the digital components of an e-mail record differs from the intellectual one,[67] it is vital to devise solutions, as simple as they may be, to appropriately describe both – for example, a register, in the form of a relational database or a spreadsheet, capable of reporting discrepancies between the two orders, could be a practical way to meet such a need.[68]

[66] Cf. 16.
[67] Cf. 18.
[68] For instance, the register might include a field where one can find the indication of the digital component that is the central intellectual component of the e-mail record (i.e., the component with relation to which all the other digital components can be regarded as its attachments). For e-mails associated with more records, another field of the register database could be intended to

3.3 MAINTENANCE AND WORKFLOWS

During and after the inclusion of an e-mail record in the RKS, the monitoring of all the actions performed on the record is needed. Audit trails must be kept both for the e-mail record as a whole and for every digital component that is a part of the e-mail record (attachments, stylesheets, digital signatures, etc.).

3.3.1 *The RKS must keep track of any event affecting an e-mail record included in the RKS. To this purpose, appropriate audit trails must be kept. The audit trail of an e-mail record must be able to indicate which digital components of the e-mail have been affected by the event. An audit trail of each digital component of the e-mail record must be also kept. The record profile of an e-mail record must show the audit trails available for the record and for its digital components. The audit trails must be considered, to all intents and purposes, as a part of the metadata of the e-mail record.*

Pennock has observed that audit trails can also "constitute an authenticity requirement".[69]

As already said above,[70] a specificity of the e-mail technology is that e-mails, and consequently e-mail records, can be made up of many digital components. Each component plays a particular role in the conceptual structure of the record. To this purpose, requirement 3.2.2 states that each e-mail record must be captured and filed so that it may distinctly show the intellectual relationships of all the digital components of which it is composed and keep track of their physical order in its metadata. To support the evidence of the integrity of an e-mail record, it is necessary to create more checksums[71] when capturing an e-mail record: a checksum of the whole e-mail record and a checksum for each of its digital components. All these checksums must be stored as metadata. If for any reason, even only one of the components is modified (e.g., owing to a process of migration involving one or more components of the record), the RKS must produce a new checksum of the e-mail record and a new checksum of each digital component that has been changed or added.

3.3.2 *Whenever an e-mail record is captured, the RKS must create a checksum of the whole e-mail record and a checksum for each of its components. All these checksums must be stored as metadata.*

3.3.3 *Whenever even only a component of an e-mail record is modified, the RKS must create a new checksum of the e-mail record and a new checksum for any of its components that*

---

contain a short description of the digital components (even just the name of the files) of the e-mail that are physically but not intellectually linked with the e-mail record.

[69] M. Pennock, *DCC. Digital Curation Manual. Instalment on 'Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages'*, Digital Curation Centre ed. (Bath: University of Bath 2006), 20. http://eprints.erpanet.org/113/01/curating-e-mails.pdf (last accessed March 28, 2010).

[70] Cf. 2.

[71] "Checksum" is a term that refers to a number of different functions. Any checksum can be defined as a hash value produced by a hash function, but different types of checksum offer different levels of protection. Cryptographic hash functions are the form of checksum that provide the highest security standards and are the only ones that are able to effectively guarantee integrity against a malicious, intentional attack. Of course, a corporation should choose the kind of checksum that best suits its business needs.

> *has been changed or modified. All these checksums must be stored as metadata in addition to the old checksums.*

E-mail records, after being integrated in the RKS, are usually used by the organization to carry out its business activities. Some rules must be enforced so the subsequent utilization complies with good records management practices. A common task in any RKS is to reproduce analog or digital records to prepare copies intended to become other records or enclosures of other records. Thanks to thread functions, with regard to e-mails, this operation can be executed by extracting an e-mail by the RKS and carrying it back in an application environment that enables the "reply" and "forward" commands. Such functions are very popular since they allow users to rapidly and flexibly communicate with other people. However, such flexibility can be counterproductive if applied without restraints in work environments where records are created. As matter of fact, when replying or forwarding an e-mail, a user can for instance, easily wipe out any trace of the headers of the former e-mail or scatter its contents throughout the new e-mail, so that the identification of the content, structure or metadata of the former e-mail becomes impossible. Moreover, a thread can be composed of tens or hundreds of e-mails, the contents of which, in theory, can be mixed because a single e-mail of the thread is able to contain the entire chain of replies or forwarded texts. In practice, it is not uncommon to see e-mails that contain, normally in the plain text part of the message, sections of four or five former e-mails that have merged together, often in such a way that it proves to be hard or impossible even to correctly identify the text of each message. Although all the e-mail clients and webmail applications are equipped with functionalities to group all the e-mails belonging to the same thread, it goes without saying that records may not be produced by irregularly amalgamating contents, structure and headers of an indeterminate number of e-mails. As Pennock rightly remarks:

> It must be clear who said what to whom. This can be difficult when people reply to a group e-mail and include the original text in their reply but insert comments in several points in the original text rather than inserting their reply as a block of text.[72]

Since in most cases the staff of an organization resort to traditional inline replies and forwarding, strict rules must be specified to direct corporate personnel to create e-mails where all the contents, structures and headers that belong to the replied or forwarded e-mail record can be effortlessly and authoritatively identified. Furthermore, the replied or forwarded e-mail record, regardless of whether it is being sent inline or in attachment, must display the metadata that describe its archival bond (classification code, unique identifier, etc.) in the RKS. Of course, all the considerations stated above are relevant to outgoing and internal e-mail records. As for incoming records, the RKS staff must take measures to prevent (insofar as it is possible) e-mail records containing parts of other e-mail messages that are difficult or impossible to be identified

[72] M. Pennock, *DCC. Digital Curation Manual. Instalment on 'Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages'*, 19.

from being integrated in the RKS, otherwise serious misunderstandings about the identity of the juridical or physical persons involved in the process of creation of the content of the record might arise. It would be strongly advisable to define guidelines and policies and again resort to the design and control of workflow processes to try to ensure that incoming e-mail records are not created through the improper use of the forward and reply functions.

We can therefore articulate the following recommendations:

3.3.4 *The RKS personnel must set rules with regard to the use of the forward and reply functions to create outgoing and internal e-mail records.*

3.3.5 *If the RKS staff notice that an incoming e-mail record contains parts of other e-mail messages difficult or impossible to be correctly identified, the RKS staff must annotate that in the record profile.*

They must restrict themselves to annotating the record profile unless, of course, a law or a relevant policy approved by the highest hierarchical level of the organization prevents such an e-mail record from being integrated in the RKS.

3.3.6 *Appropriate policies and business process workflows should be set to try to avoid that the incoming e-mails may be created through an improper use of the forward and reply functions.*

3.3.7 *Regardless of any association existing with a thread of e-mails, any e-mail record must always be included in the RKS in a way that mirrors its intellectual relationships with all the other records of the RKS and the whole of the records kept in the RKS. All the information concerning its association with the thread to which it belongs must be stored in its metadata.*

Within an RKS, the treatment of e-mail records does not differ from that of other digital records with regard to inclusion in corporate workflows and possible duplications. A positive side-effect of the integration of e-mail records in an RKS is the end of the production of unnecessary copies, which otherwise have to be sifted through to identify corporate records.[73] In RKSs, a digital record is not physically reproduced if it relates to two different transactions: its metadata are changed so that it is linked with two different files or classes of a classification plan and, from an intellectual point of view, the actual result of such a change is the presence of two digital records in the RKS.

If the step of a workflow provides that an office within the organization or a member of the corporate staff must be formally informed about an e-mail record newly included in the RKS, of course the RKS must somehow keep track of the fact that the communication has taken place (for example, by modifying the record profile of the e-mail record).

[73] Cf. New Zealand Archives, "Fact Sheet E-mail", 2006, 1. http://continuum.archives.govt.nz/files/file/factsheets/f10.pdf (last accessed March 29, 2010).

In this respect, e-mails feature the well-known return-receipt function, which is intended to alert the sender if the addressee has received the message: such a function could be construed as a very simple workflow tool. Nonetheless, return receipts are unreliable as a means to handle workflows, both because the addressee can refuse to send the notification and because they, in their basic form, certify only that an e-mail has been opened in an account, which does not means that a given person has actually read the e-mail. However, from the diplomatics perspective, the return receipt function creates a notification e-mail, which has to be captured according to the procedure in force for all the other e-mail records to be included in the RKS. A notification e-mail record must be filed together with the e-mail record to which it relates, but is not an attachment or a component of that record, whose metadata must be updated, if the notification of the e-mail record to a given account is an essential step to be logged.

3.4 LONG-TERM PRESERVATION FORMATS AND SOLUTIONS

Long-term preservation of e-mail records is particularly complicated because an e-mail may be composed of components encoded in different formats. This raises problems even in the active phase of the record life cycle, since the format of one component of an incoming e-mail record could be unreadable in a given RKS. For long-term preservation, prospective troubles are even greater, since records managers and archivists have to figure out how to guarantee in the course of time the authenticity, reliability and usability of records that can be extremely multifaceted. All the research projects that have dealt with this issue (e.g., DAVID, Digitale Duurzaamheid, TERM,[74] NHPRC[75] Funded Preservation of Electronic Mail Collaboration Initiative, the design and implementation of the XENA software by the National Archives of Australia) agree that the action for long-term preservation must be started in conjunction with the beginning of the life cycle of e-mail records (as the authors of an article about the TERM Project correctly note, "long-term preservation of electronic records begins with the creation of the e-mail messages themselves"[76]).

The use of XML is a staple of these research projects. All the scholars who have committed themselves to studying this problem contend that this standardized syntax possesses many features that make it suitable for long-preservation purposes. First of all, XML is particularly fit to describe the structures and properties of textual documents, and e-mail technology has been developing just from specifications designed for plain text messages encoded in ASCII – even today, when we think of an archetypal e-mail message, we usually picture in our mind a textual document in conjunction of some text headers. Second, languages created by means of XML syntax are suitable for hierarchical representations of the information of a document, a characteristic that turns out to be very useful to express both the internal

[74] Texas Email Repository Model
[75] This project, funded by the NHPRC (National Historical Publications and Records Commission), has been furthered by the North Carolina State Archives, the Kentucky Department of Library and Archives, and the Pennsylvania State Archives.
[76] P. Galloway, M. Green, S. Gunn and S. Soy. "Coming to TERM Designing the Texas Email Repository Model," in *D-Lib Magazine* 8, no.9 (September 2002). http://www.dlib.org/dlib/september02/galloway/09galloway.html (last accessed March 29, 2010).

organization of entities made up of many parts, the order of which "is meaningful",[77] and metadata that can refer to a whole entity or only to some of its parts. Besides, thanks to the ever-growing production of new XML-based standards and languages, a greater and greater range of data and document types can be encoded in XML. These characteristics must be added to the advantages of XML as a method of representation independent of specific technological tools and implementations and therefore suitable for supporting format conversion processes and portability across time. In addition, XML is also both human and machine readable and not subject to copyright restrictions, because it is an open and public standard.

Consequently, it is not surprising that almost all the research projects for the long-term preservation of e-mails include XML as one of the key elements of the adopted strategy. XML representations are often used to prepare OAIS-compliant AIPs (Archival Information Packages), which wrap an e-mail, including all its components, together with its metadata, so that each e-mail is in a sense joined with the information intended to describe it. An e-mail stored in this kind of AIP has been exported in a format for long-term preservation, while the AIP can also contain other formats of the e-mail or the original bitstream, that is to say, the bitstream of the e-mail, when it was initially created. The final result of such an operation can be a rather complex object, comprising several different entities. The researchers of the project "Digitale Duurzaamheid" have given an example of a list of items that could be included in one of these packages:[78]

- The message text (body), if possible in more than one form (for example, in both a plain text layout and in HTML)
- Attachments: images or other objects (often present in e-mail messages marked up in HTML) that belong with the e-mail message
- Metadata about context

And in addition

- A preservation log file (in fact an audit trail of the actions carried out on the e-mail message, and the technical data needed for preservation)
- The original transmission file, and
- Extra metadata

An approach that aims to build objects that contain a rich array of metadata, formats, entities, information about structures and contexts, wrap and if applicable, even encode them in languages based on a widespread and open standard syntax, which is also not bound to specific technological implementations, is certainly the right one. Moreover, XML can also be exploited to further records management. We have already seen[79] that XML Schemas can be used to create templates and predefined documentary forms that can interact with RKSs and any IT system,

[77] G. Pontevolpe and S. Salza, 17.
[78] Digital Preservation Testbed. *From digital volatility to digital permanence. Preserving email*, 69.
[79] Cf. 20.

and automatically trigger a number of events. Nevertheless, in spite of all these positive factors, there remain some problems. XML is an extremely powerful tool to represent textual data, but obviously is not able to encode everything, such as audio and video attachments, executable files and raster images. Apart from that, it is unrealistic to demand that people completely give up formats that are deeply rooted in ordinary usage, such as PDF and MS Access, or other proprietary formats designed for special applications. Therefore, it is necessary to devise a plan providing operations of conversion and/or the coexistence of more formats. In this respect, stylesheets (based on the XSLT language) are a flexible and powerful instrument, but cannot solve every problem and what is more, an organization must also take into consideration copyright concerns when a digital conversion program is being enacted.

With regard to e-mails, there is a third aspect to be carefully evaluated. Even after the inclusion in an RKS, corporate staff could need to transfer an e-mail record back to the client or webmail application used to handle e-mails before its capture in the RKS, for instance to consult it or attach it to another e-mail. In this case, if an e-mail record has already been transformed in a preservation object encoded and wrapped in XML together with many other digital items, it will prove to be impossible to re-use it. The issue of compatibility between long-term preservation solutions and e-mail client software led the scholars involved in the DAVID Project, who had set up a procedure to export the e-mails of the Antwerp city archives from a client environment to a records management application, to resort to a format that could be reopened in the e-mail client, to expedite the daily work of the staff. E-mail records were migrated to a format suitable for long-term preservation only when accessioned into the digital repository.[80]

DAVID researchers have rightly allowed for the need to ensure the effective orchestration of the business activities of the corporation chosen as a testbed for their project. There is no point in conceiving organizational schemes and solutions that are seemingly perfect, but burden beyond reason the available resources of an organization or are impossible to be implemented for any other reason. Any records management program must be feasible and viable to be supported by the organization that should benefit from it. This can be explained using OAIS terminology: if the process of integration of an e-mail record in an RKS creates an AIP, records managers and archivists also have to think how to obtain from that AIP a DIP (Dissemination Information Package) that is able to be of use to the corporate staff. The option to delay the migration to long-term preservation formats until the end of the active phase of the record life cycle should be only a measure of last resort. This is due to the fact that e-mails often include attachments that feature proprietary or uncommon file extensions, the treatment of which for preservation purposes must be executed as soon as possible, since relevant documentation and processing tools could become quickly unavailable. To duplicate each e-mail record, by creating one to be kept in native format and another one to be preserved in an OAIS-compliant AIP, is dangerous and entails the risk of misalignment. In fact, some maintenance interventions could be required to guarantee the ongoing serviceability of one of the two or both

[80] F. Boudrez. *Filing and Archiving E-Mail*, 19-20.

records (for instance, subsequent migrations of AIPs; additions of timestamps to digital signatures).

Pennock suggests that one produces on-the-fly copies for consultation purposes to – as she states – "protect the integrity of the stored records whilst still providing users with freedom to manipulate retrieved records."[81] If we try to flesh out such a piece of advice, we could say that a DIP could be extracted on the fly from the AIP kept in the RKS.

We have to remember that we cannot assume anything about the way to build an AIP: OAIS is an abstract model and an AIP can be a single multifaceted object wrapped in XML as well as a series of objects that are physically separated but logically related. As B.F. Lavoie correctly remarks:

> The archived information and its associated metadata represent a single logical package within the archival system: there is, however, no requirement that any form of physical association be maintained.[82]

If we want to stick to a kind of implementation of an AIP realized by means of an XML wrapping that physically encapsulates all the components of an e-mail record, we could build an entity including two formats for each component: the same format in which the component was encoded to be handled by the e-mail client and another one designed for long-term preservation. In reference to what has already been stated in section 3.1,[83] it is logical to argue that a practical and useful way to put into effect recommendation 3.1.4, which prescribes to do one's best to integrate e-mail records in the RKS, might be to develop plug-ins that enable the RKS both to receive e-mails exported from the e-mail client and to generate instantly from an AIP a DIP that can be opened by the e-mail client or another application. To set up such add-ins should be a task within the reach of many organizations, since what is needed could be a breakdown of an XML object (assuming that an AIP has been framed with XML), especially if one of the components of an AIP is a version of the e-mail record in a format that can be processed by the e-mail client.

Nevertheless, we do not need to generate AIPs that materially embed all the components of an e-mail record. We could keep sets of objects lying in different locations, but logically linked through appropriate metadata elements, which include the global checksum of the whole AIP and all the checksums of its single components (cf. above recommendations 3.3.2 and 3.3.3). If one of the checksums of the AIP should be modified, the RKS software would immediately warn the staff. In such a frame of reference it would be far more straightforward to include in the AIP more formats of the same component and then, if need be, extract one of

[81] M. Pennock. "Managing and Preserving E-mails," UKOLN, 2006, 5. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/RMS-b_mngmt-pres-emails.pdf (last accessed March 30, 2010).

[82] B.F. Lavoie. "The Open Archival Information System Reference Model: Introductory Guide," DPC Technology Watch Series Report 04-01 (OCLC and DPC 2004), 11. http://www.dpconline.org/component/docman/doc_download/91- introduction-to-oais-introduction-to-oais (last accessed June 2, 2011).

[83] Cf. 15.

these together with a subset of metadata to produce a DIP for consultation or use in a given application environment. In conclusion, there are several options that can accommodate effective records management practices with the immediate conversion of e-mail records to long-term preservation formats.

Usually, however, the most serious difficulties with regard to procedures of exportation and migration are brought about by attachments. To this purpose, recommendation 3.2.23 establishes that the RKS personnel have to implement a list of the formats in which the digital components of e-mail records may be encoded, since it cannot be ruled out that some formats are not even suitable for treatment by the RKS in the course of routine records management practices. Obviously, such a list must include the formats that the RKS is able to manage, but should also be set up in coordination with BPR activities, so as to detail which formats are required to carry out some significant corporate transactions and which metadata elements must be available to adequately document them. The database should also contain other relevant pieces of information, such as the long-term preservation format to which a file with a given extension must be converted for permanent retention purposes. As recommendation 3.2.23 prescribes, the list of formats also has to be consistent with the retention and disposition schedules established in an organization; if an e-mail record belongs to a class of records selected for permanent retention, the requirements about the formats of the digital components of that e-mail must be stricter. The RKS should be linked with such a list to be provided with all the information needed to assess whether a format associated with a specific e-mail record created in the course of a given transaction is suitable and if, in the case in point, all the conditions required for a correct procedure of integration in the RKS have been fulfilled. The result of this assessment should be available to the user being accountable for the operation of inclusion. To give some examples, an e-mail containing an HTML component that is not XHTML-compliant could be captured only after the component has been converted to XHTML; an organization may establish that, if an applet java is appended to an e-mail, the e-mail may be captured only if the source code of the applet java is available.

It is easy to infer from what has been said above that migration and extensive use of standards are suggested as the key elements of a strategy for long-term preservation. Emulation is not a viable option since the digital components of e-mail records are too diversified to be preserved by resorting to such an approach. Furthermore, as Potter has astutely remarked, "the sender and receiver will often perceive the look and feel attributes of a message differently."[84] E-mail technology is based on standard exchange formats aimed at enabling different systems (i.e., different e-mail clients and webmail applications) to cooperate with one another, regardless of application-specific visual representations of contents and structure, so that it is not clear what an emulation-based process should target.

[84] M. Potter. "Researching Long Term Digital Preservation Approaches in the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring)," in *RLGDigiNews* 6, no.3 (June 15 2002). http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2974.html (last accessed March 31, 2010).

On the basis of all the previous considerations, the following recommendations can then be articulated:

3.4.1 *The solutions and procedures aimed to ensure the long-term preservation of e-mail records must be developed within a framework of strategies based on migration and use of standard formats and software- and hardware-independent methods of representation of information.*

3.4.2 *Regardless of any specific solution adopted to further the long-term preservation of e-mail records, the RKS must preserve all the digital components of e-mail records in formats that allow to efficiently and effectively carry out the routine business activities of the organization.*

3.4.3 *The organization's staff should use only formats suitable for long-term preservation to create outgoing or internal e-mail records designed for permanent retention.*

3.4.4 *The RKS staff should try to enforce the use of formats suitable for long-term preservation for incoming e-mail records designed for permanent retention by resorting to policies and business process modelling.*

3.4.5 *When an e-mail record designed for permanent retention is captured into the RKS, the RKS should convert all the digital components of the e-mail record encoded in formats that are not suitable for permanent retention to formats fit for long-term preservation purposes.*

3.4.6 *When an e-mail record designed for permanent retention is captured into the RKS, the RKS should link together all the digital components of the e-mail record, all the metadata describing them and other possible digital components and metadata added to the e-mail record for long-term preservation purposes in an OAIS-compliant AIP.*

The production of well-designed AIPs can be very useful also for current records management, besides promoting long-term preservation, because the information contained in an OAIS-compliant AIP can be used by an RKS to execute routine records management operations and support business workflows. However, the resources necessary to establish OAIS-compliant AIPs even in the active stage of the records lifecycle could be too demanding for organizations that do not need to face long-term preservation issues or cannot afford a process of permanent retention without the support of external agents, so that the effort required to create OAIS-compliant AIPs can be considered essential only to pursue long-term preservation purposes. Therefore, recommendation 3.4.6 is only advisable.

3.4.7 *When an e-mail record that must be permanently retained is moved from the RKS to the digital repository, the RKS must convert all the digital components of the e-mail record encoded in formats that are not suitable for permanent retention to formats fit for long-term preservation purposes. The RKS staff must ensure that the RKS preserves,*

throughout all the active phase of the life cycle of the e-mail record, all the metadata and the resources necessary to correctly perform the conversion.

Recommendation 3.4.7 is one of last resort. If conversion to long-term preservation formats is delayed until the accession to the digital repository, the RKS staff risk being no longer able to carry out the process of conversion. Therefore, adequate measures must be taken to prevent that from occurring.

3.4.8 *When an e-mail record designed for permanent retention is moved from the RKS to the digital repository, all the digital components of the e-mail record, all the metadata describing them and other digital components and metadata added to the e-mail record for long-term preservation purposes must be logically linked together in an OAIS-compliant AIP.*

Apart from being the framework of all the main research projects on e-mail long-term preservation, OAIS has been termed by the authors of the Preservation Task Force Report of InterPARES 1 as "the basis for the content of the preservation process model"[85] developed by them. However, one must keep in mind that OAIS is a high-level model, so that on the one hand it can be easily adapted to arrange a long-term preservation plan that meets the needs of a given environment, while on the other hand the quality of the outcomes strongly depends on the appropriateness of the chosen solutions. Because OAIS prescribes no specific tool or metadata, the records managers and archivists of an organization must be capable of understanding which pieces of information have to be added to an AIP.

Obviously, the integration of e-mail records in the RKS also implies that the strategy for e-mail records long-term preservation must be coordinated with that established for all the other corporate digital records selected for permanent retention. Broadly speaking, the metadata elements to be included in e-mail record AIPs are not different from those required for other categories of digital records: what distinguishes e-mail records is above all the set of attributes describing the physical configuration and the intellectual structure of e-mails and the operation of transmission. Such attributes can be effectively captured only when e-mail records are included in the RKS, which once again shows that the organization of a proper e-mail capture process is of paramount importance. Among metadata elements, those related to the custodial history are particularly significant and clearly can never be omitted because they convey essential elements to reconstruct the history and corroborate the trustworthiness of any AIP. In fact AIPs, and especially highly-structured AIPs, as those of e-mail records often are, need regular maintenance: new migrations (which in turn entail the generation of new metadata), inclusions of new metadata, upkeep of AIP audit trails. These latter observations can give grounds for two more recommendations:

[85] InterPARES 1, "Preservation Task Force Report," 2001, 9. http://www.interpares.org/book/interpares_book_f_part3.pdf (last accessed April 1, 2010).

3.4.9 *The RKS staff must set up an appropriate procedure to ensure the ongoing maintenance of e-mail record OAIS-compliant AIPs and schedule all the operations required to continuously guarantee their usability and trustworthiness.*

3.4.10 *All the metadata documenting the custodial history and maintenance of an e-mail record OAIS-compliant AIP must be added to the AIP.*

Of course, such recommendations are also valid for other types of digital records.

3.5 LINKS, DIGITAL SIGNATURES AND HIDING ENCRYPTIONS

Some common features of e-mails may give rise to serious problems when an e-mail is associated with one or more records.

E-mails are self-contained entities, but as said in section 2.1,[86] links embedded in an e-mail record can point at external resources that are outside the control of the RKS. Such a situation can be difficult to handle if an external resource is indispensable to avail oneself of the e-mail record to carry out a transaction. For instance, a link can be placed in an e-mail in substitution of an attachment. In the strict meaning of the term, these resources are not parts of the record; nevertheless, their role can actually be far more important than that of references in analog records. The RKS staff must always adequately describe the external resource and must be able to help the organization to specify in which cases it is inappropriate that determined elements essential to use an e-mail record created in the course of a given transaction are kept in an external resource and, in the end, must try to prevent such cases by resorting to appropriate policies and business process modelling activities. The reach of the expression 'adequately describe' must be defined by each organization in relation to the legislative and regulatory framework and its business needs, but the description must include at least the indication of the URI, the outcome of a check test on the links (to verify that they are running), the date and time of the check test and, whenever possible, the juridical or physical person accountable for the external resource. If an external resource contains elements indispensable to use the e-mail record to carry out a transaction, the RKS staff should try to obtain any other kind of metadata that turn out to be useful to give evidence of the configuration of such a resource at the time of the capture of the record in the RKS: for example, if applicable, a checksum of the external resource (it could be sent by the person accountable for the external resource).

We can then formulate two simple recommendations:

3.5.1 *At the time of the capture of an e-mail record, the RKS staff must appropriately describe every link connected with an external resource. The description must at least include the URI of the resource, the outcome of a check test on the link, the date and time of the check test and the juridical or physical person accountable for the external resource. If necessary, the description must also include any other kind of metadata that turn out to*

[86] Cf. 8-9.

be useful to give evidence of the configuration of such a resource at the time of the capture of the record in the RKS.

3.5.2 *The RKS staff must be able to help the organization to specify in which cases it is inappropriate that determined elements essential to use an e-mail record created in the course of a given transaction are kept in an external resource, and must try to prevent such cases by resorting to suitable policies and business process modelling activities.*

As always, it is easier to enforce policies and rules in relation to outgoing and internal e-mail records, because it is simpler to require a member of the organization to create an e-mail record that complies with given rules.

Digital signatures are often appended to an e-mail record to identify the person who has signed it. Digital signatures by themselves are not sufficient to establish the authenticity of a record, since they certify the integrity of a bitstream at a given time (while records can consist of more bitstreams)[87] and that a person in possession of a private key has signed the bitstream. However, a digital signature is a diplomatic element of a record and the traces of it (i.e., appropriate descriptive metadata) must be preserved throughout the retention period scheduled for that record. Since digital signatures are applied to a bitstream, records managers have to pursue two goals in the same time: 1) to protect the integrity of the bitstream and the evidentiary value of the digital signature for the purpose of attesting to such integrity; 2) to preserve the digital signature as a diplomatic element of the record associated with the bitstream, to support the authenticity and reliability of the record. These two objectives are different and often clash with each other. In principle, the demonstration of the validity of a digital signature requires, besides the integrity of the bitstream, a series of objects such as the PKI (Public Key Infrastructure) certificate, the related timestamps, the hashing algorithm, the tools to decrypt the digital signature and to access and use the original bitstream, as well as the metadata related to these objects. The maintenance of all these items requires the implementation of a validation chain, which, in a sense, serves as the chain of preservation of records, although the meaning and context of the two chains are completely dissimilar from each other. Digital signatures must also be taken into consideration in relation to long-term preservation of records; however, the main function of digital signatures is to demonstrate the trustworthiness of records, not the integrity of bitstreams and the identification of private keys. The authenticity of records is an important element of their trustworthiness, but to this purpose, what matters is to thoroughly describe them and to give evidence of the whole process of preservation from the point of their creation. In a sense, an authentic record in turn can attest to the authenticity of a digital signature appended to it (or of the metadata describing a digital signature, the bitstream of which has been destroyed owing to a migration process), in the same way as the full documentation of the circumstances and context of an investigation and of the ensuing trial can support the authenticity of a DNA profiling carried out in the course of a legal action. The insightful

[87] F. Boudrez. "Digital Signatures and Electronic Records," in *Archival Science* 7 (2007), 183.

comparison between the evidentiary value of a digital signature and that of a DNA profiling must be credited to the four authors of a report by the LongRec (Long-Term records Management) Project, a three-year project (2007-2009), which was developed by the Norwegian team of InterPARES 3. Groven, Ølnes, Abie and Fretland, through an analysis of the events that occurred in 1995 during the well-known O.J. Simpson trial, have cleverly shown how no proof is completely freestanding, because it must be in turn certified by the traces of the transactions and procedures that led to its discovery or production, and correctly conclude that:

> It is those traces that were successfully contested during the Simpson trial, because, as archivists have long known, no evidence is ever self-intelligible[88]

Therefore, records managers and then archivists have to handle digital signatures through two differentiated approaches, which are developed by taking into account the framework of laws, regulations and policies that affect each record. E-mail records that have been selected for permanent retention require that the action aimed to guarantee the validity of a digital signature (and based on the integrity of the bitstream and the maintenance of the entire validation chain) has to be accommodated with a more global long-term preservation strategy. This strategy entails migration and using contextual metadata, instead of validation chains, to document the validity of the digital signature at the time that it was appended to the record and throughout the period during which the signature remained functional. In practice, that means the original bitstream must be preserved as long as the business need to maintain the validity of a signature persists and to be able to perform all the operations that relate to the validation chain. It is foreseeable that sooner or later such a need will cease and then it will be no longer necessary to retain the original bitstream, which will be able to be converted to a long-term preservation format. If one develops an action of long-term preservation of a e-mail record associated with a digital signature as soon as the record has been included in the RKS, this implies that two formats of the digital component to which the digital signature has been appended would be kept. The first one is that of the original bitstream of the component combined with the valid digital signature (of course, such a format must be conserved only as long as the validity of the signature is required because of legal obligations and business needs), the second one is the format designed for long-term preservation. An OAIS-compliant AIP can be adapted to logically or even physically link (for instance through a XML encapsulation) the two formats: with regard to that, the authors of the LongRec report have written that:

[88] A.-K. Groven, J. Ølnes, H. Abie and T. Fretland. *Preservation of Trust in Long-Term Records Management Systems. A State of Art Overview for the LongRec Project* (Oslo: Norwegian Computing Center 2008), 27. http://research.dnv.com/longrec/Intranet/ResearchResults/StateOfTheArt/longrec-trust-soa-report-final.doc (last accessed June 2, 2011).

> Encapsulating both the original digital content (bit stream) and associated metadata, together with (all) derivations of content and associated metadata, seems to be fruitful.[89]

This approach can be enacted if an organization has enough resources to pursue it, otherwise only the original bitstream can be kept, as long as the digital signature serves to carry out the routine business activities of the organization (cf. recommendation 3.4.2). However, it is apparent that in the long run the preservation of the validation chain of the digital signature will become untenable and also useless. In consideration of all that has been said, the following recommendations can be articulated:

3.5.3 *The RKS staff must ensure the validity of any digital signature appended to a digital component of an e-mail record by preserving the original bitstream of the digital component and performing all the actions required to operate the validation chain of the digital signature, as long as the validity of the signature must be retained because of legal obligations and business needs. The RKS staff must register in the record profile the metadata needed to appropriately describe the digital signature.*

3.5.4 *If a digital component is associated with a digital signature and the e-mail record to which the digital component belongs has been selected for permanent retention, the RKS staff should convert the digital component to a format fit for long-term preservation and to be kept together with the original bitstreams of the component and the digital signature as soon as the e-mail record has been included in the RKS, and register in the record profile the metadata needed to appropriately describe the traces of the digital signature.*

3.5.5 *If a digital component is associated with a digital signature and the e-mail record to which the digital component belongs has been selected for permanent retention, the RKS staff must convert the digital component to a format fit for long-term preservation and to be kept together with the original bitstreams of the component and the digital signature as soon as the e-mail record has been moved from the RKS to the digital repository and the digital signature is no longer needed to meet the legal obligations and business needs by reason of which it has been created, and register in the record profile the metadata needed to appropriately describe the traces of the digital signature.*

3.5.6 *As long as the digital signature of a digital component of an e-mail record continues to be valid, the RKS staff must register in the record profile the metadata needed to appropriately describe any action and modification performed to operate the validation chain of the signature.*

Recommendation 3.5.6 is useful not only for long-term preservation purposes, but also to strengthen the evidentiary value of the digital signature during the time that it serves to meet the

[89] A.-K. Groven, J. Ølnes, H. Abie and T. Fretland, 47.

legal obligations and business needs by reason of which it has been created, because such metadata attest to all that has been carried out to operate the validation chain.

The digital signature is a form of encryption used to detect whether a bitstream has been tampered with. There are other kinds of encryptions, the purpose of which is not to demonstrate the integrity of a bitstream, but to hide the information contained in a record (e.g., sensitive personal information).[90] As in the case of a digital signature, the original bitstream of a hiding encryption must be kept as long as the encryption serves to meet the legal obligations and business needs by reason of which it has been created, so that, in principle, hiding encryptions do not differ from digital signatures with regard to the general guidelines to be observed by the RKS and digital repository staff. Therefore, the recommendations for hiding encryptions are not different from those suggested for digital signatures and a possible action for long-term preservation must be coordinated with the constraints imposed by the presence of a hiding encryption. However, as a rule, the validation chain of a hiding encryption is less complex than that of a digital signature, since its creation and maintenance does not require so many resources and procedures as those one has to set up to run a PKI-based system.[91]

4. CONCLUSION

E-mails are a fundamental component of the ever-increasing amount of records and documents that often overwhelms corporations and are a side-effect of the development of systems that make it far easier to produce and transmit documentary traces of one's own activities.

From a records management perspective, the key factor to be able to properly manage the huge mass of e-mails lies in the ability to set up an environment that is both targeted to the requirements the records creator has to comply with and flexible enough so as to be modified as appropriate to meet new possible needs. To that end, it is necessary to build a thorough understanding of the business processes and workflows of the records creator to optimize and automate records management procedures as much as possible without losing any information essential for an adequate description and treatment of the e-mail records, also with reference to long-term preservation issues. To achieve such an understanding involves an initial investment of time and effort and the constant monitoring of the changes affecting the work environment of the records creator, but gives in return remarkable efficiency improvement and time savings and is the only strategy through which a corporation can effectively cope with more and more substantial and complex flows of e-mail records.

[90] E.g., Massachusetts has established that any sensitive personal information is to be encrypted when it is transmitted electronically or stored on portable devices (201 CMR 17.00: Mass. Gen. Law c. 93H, effective January 1, 2009). Cf. C.
[91] C. Komanecki. "The Move Toward Mandatory Encryption of Sensitive Personal Information," World Law Group, 2. http://www.theworldlawgroup.com/files/file/docs/Faegre%20Encryption.pdf (last accessed May 12, 2011).

The set of recommendations specified in the former section implies an ongoing analysis of the business processes and workflows to identify the most suitable solution when a recommendation is to be applied to a particular case and avoid a waste of resources.

Even before the appearance of digital records, the full knowledge of how the corporation carried out its activities `and transactions was a basic skill of any records manager because it provided him or her with indispensable information for tasks such as drafting a classification plan, determining retention and disposition schedules and granting access rights. Nowadays, such knowledge is the only way that may enable record managers to play a proactive role in their workplaces and design systems where records management practices are integrated in the overall business context and therefore contribute to create value and fulfill the mandate of the corporation.

APPENDIX 1

LIST OF THE RECOMMENDATIONS AND ASSOCIATED RISKS OF NON-COMPLIANCE

LEGEND: E = Essential A = Advisable

CODE	RECOMMENDATION	DEGREE OF COMPLIANCE		NON-COMPLIANCE RISK(S)
3.1.1	*All the e-mail records made or received, and kept by an organization must be integrated in the RKS of the organization.*	E		This is the pivotal recommendation, which is the indispensable basis for any viable management system for e-mail records. The recommendation is valid even for physical persons. At least from an intellectual point of view, integration must be always pursued. Not abiding by this recommendation automatically means undermining any serious effort to establish effective e-mail recordkeeping practices.
3.1.2	*In an organization business activities must be carried out by using as much as possible corporate accounts, which have been associated with particular offices or have been established to fulfill specific functions, and can be freely accessed at least by given members of the RKS staff.*	E		Failure to comply with this recommendation may lead to loss of control over e-mail flows and give rise to privacy issues if the RKS staff need to access individual business accounts.
3.1.3	*An organization must draw up a list of the e-mail accounts enabled to directly interact with the RKS.*	E		This recommendation is strictly connected with the former one. Non-compliance makes it difficult to avoid loss of control over e-mail flows. Compliance should also reduce the risk of privacy issues.
3.1.4	*An organization must select tools and strategies intended to facilitate the integration of e-mails records in the RKS.*	E		In other words, proactive behaviour is anyway needed: the pervasiveness and constant flow of e-mails make it impossible to devise effective makeshift solutions, suitable for coping with both daily business operations and emergency situations.

3.2.1	*The operation of capture and filing must be performed only through the e-mail accounts enabled to directly communicate with the RKS.*	E		This recommendation is a direct corollary of recommendation 3.1.3. Non-compliance entails the serious risk of unrestrained proliferation of incoming and outgoing e-mail flows.
3.2.2	*The operation of capture and filing must include e-mail records in the RKS so as to always highlight their intellectual structure. The information about the physical order of the digital components that make up an e-mail record must be always preserved as metadata in the record profile of the e-mail record.*	E		In other words, it is necessary to keep track of both the intellectual and physical structure of an e-mail record. Failure to comply with this recommendation automatically makes it impossible to appropriately manage e-mail records and leads to the loss of essential metadata.
3.2.3	*The process of capture and filing of e-mail records in a RKS should be automated as much as possible.*		A	This recommendation is advisable, since there are also small organizations that might not need any automated procedure. However, for large organizations it is unthinkable to do without a machine-aided approach: in most cases the sheer volume of e-mails would disrupt their recordkeeping system.
3.2.4	*In cases where the process of capture and filing is partly automated, the RKS staff must implement procedures that let humans audit the work carried out by machines and guarantee accountability for the decisions that have been made.*	E		No machine can be held accountable for its work: in case of litigation, the absence of such procedures involves huge risks. Apart from that, absolute reliance on automation may prove to be extremely dangerous also with regard to the development of business activities: humans must always have the last word.
3.2.5	*Templates and recurrent patterns intended to design regular documentary forms to be applied to e-mail records should be arranged by the RKS staff to expedite automated and human intellectual analysis of e-mails with a view to their capture and subsequent handling in the RKS.*		A	Management-oriented diplomatics is a fundamental element of an approach aimed at enabling organizations to cope with gigantic flows of e-mails. Non-compliance is likely to result in failure to adequately support a corporate e-mail records management plan.
3.2.6	*The RKS staff should resort to Business Process Reengineering modeling and analysis to understand which kinds of records are needed to carry out given transactions and which characteristics and sets of metadata elements and values must be present in such records.*		A	Business Process Reengineering is a fundamental element of an approach aimed at enabling organizations to cope with gigantic flows of e-mails. Non-compliance is likely to result in failure to adequately support a corporate e-mail records management plan.

3.2.7	*The RKS staff must define the list(s) of digital and intellectual features needed to make up e-mail records and the list(s) of metadata that must be associated with an e-mail record*	E		These lists are essential to understand whether an e-mail record is able to fulfill the purposes for which it has been created. Risks of non-compliance are therefore self-evident.
3.2.8	*Any record sent by e-mail must be included in the RKS as a separate entity.*	E		For the purpose of appropriately manifesting the intellectual structure of the records and of the whole RKS, each record must exist as a separate entity in the RKS. Appropriate records management is impossible in case of non-compliance.
3.2.9	*Any e-mail should be associated with one and only one record.*		A	Non-compliance with this recommendation makes it far more difficult to abide by recommendation 3.2.8, which is an essential one.
3.2.10	*The sender address of an e-mail record must be associated with an identified physical or juridical person in conjunction with the inclusion of the e-mail record to the RKS. If the association proves to be impossible, the author of the e-mail record must be regarded and registered as an unknown entity.*	E		In case of non-compliance, the identification of the persons who have taken part in the creation of the e-mail record has not been accomplished. Such identification is always an indispensable step, even if its final outcome is that the author of an e-mail record is an unknown person.
3.2.11	*The RKS staff must ensure that one or more lists (e.g., databases, spreadsheets) to link sender addresses with the physical or juridical persons to which they relate are kept and updated. The RKS must be connected with the list(s) to make the identification of a sender easier. Procedures and means employed to check the identity of the physical or juridical person associated with a given sender address must be reported in the list(s).*	E		Non-compliance makes it difficult or even impossible to set up a viable and effective procedure for the successful identification of the persons who have taken part in the creation of an e-mail record.
3.2.12	*All the e-mail accounts of the addressees of an e-mail record must be associated with an identified person in conjunction with the inclusion of the e-mail record in the RKS. If the association proves to be impossible, the addressee to which that account relates must be regarded and registered as an unknown entity.*	E		In case of non-compliance, the identification of the persons that have taken part in the creation of the e-mail record has not been accomplished. Such identification is always an indispensable step, even if its final outcome is that one or more addressees of an e-mail record are unknown persons.

3.2.13	*The RKS staff must ensure that one or more lists (e.g., databases, spreadsheets) to link the e-mail accounts of addressees with the physical or juridical persons to which they relate are kept and updated. The RKS must be connected with the list(s) to make the identification of an addressee easier. Procedures and means employed to check the identity of the physical or juridical person associated with a given e-mail account of an addressee must be reported in the list(s).*	E		Non-compliance makes it extremely difficult or even impossible to set up a viable and effective procedure for the successful identification of the persons that have taken part in the creation of an e-mail record.
3.2.14	*If a distribution list is in the 'to' header of an e-mail associated with one or more records, the RKS staff must strive to gather all the available documentation about the distribution list. The documentation concerning the distribution list must be kept in the RKS and must be reported in or linked with the record profiles to which it pertains.*	E		Distribution lists are likely to make it difficult or even impossible to properly identify all the addressees of an e-mail record. Risks of non- compliance are the same as those specified for recommendations 3.2.12 and 3.2.13.
3.2.15	*In relation to outgoing and internal e-mails, an organization must set rules about the use of the headers 'carbon copy' and 'blind carbon copy' and corporate policies must establish when an e-mail account of an addressee can be placed in the 'carbon copy' and in the 'blind carbon copy' headers.*	E		Without appropriate rules it is impossible to avoid an improper use of the 'carbon copy' and 'blind carbon copy' headers. Furthermore, blind carbon copies hide meaningful pieces of information and consequently prevent addressees and recordkeeping systems from extracting important metadata from e-mails.
3.2.16	*Recommendations 3.2.12, 3.2.13 and 3.2.14 must be applied to all the e-mail accounts and distribution lists placed in the 'carbon copy' and 'blind carbon copy' headers.*	E		Risks of non-compliance are the same as those specified for recommendations 3.2.12, 3.2.13 and 3.2.14.
3.2.17	*The RKS staff must ensure that the subject of an e-mail record to be included in the RKS is reported in the record profile and that the subject meaningfully expresses the transaction or the matter to which the record relates.*	E		The proper indication of the subject of a record is indispensable for some operations such as classification and filing, and is useful to determine especially the procedural context and the content of the record. In addition, in some countries such an indication is mandatory for public bodies. Risks of non-compliance therefore are self-evident.

3.2.18	*The RKS staff should ensure that the subject header of any outgoing and internal e-mail meaningfully expresses the subject of the e-mail record.*		A	Without a meaningful subject header, the determination of the subject of the record(s) associated with an e-mail is obviously more difficult.
3.2.19	*The RKS staff must ensure that the subject header of any incoming e-mail to be included in the RKS is captured the way it is and added to the record profile of each record associated with the e-mail. If the subject header does not meaningfully express the subject of the e-mail record or expresses it wrongly, the RKS staff must anyway ensure that the record profile reports an adequate definition of the subject of the record.*	E		Of course, the need for an appropriate indication of the subjects of records does not mean that e-mail subject headers can be altered, since these are essential metadata elements that must be always registered the way they are. It is a necessary approach to accommodate different requirements. Again, risks of non-compliance are self-evident.
3.2.20	*The RKS staff should set up appropriate tools (e-mail templates, databases of normalized subjects, controlled workflows) to expedite the definition of an appropriate subject for any e-mail record and its insertion in the associated record profile and, if possible, in the subject header of the e-mail.*		A	To determine the appropriate subject for a record always requires some intellectual analysis. If the task becomes too time-consuming, organizations and their RKS staff are likely to drop it or not to properly perform it. Adequate proactive measures are therefore needed to avoid this risk.
3.2.21	*When an e-mail record is captured by the RKS, the RKS must ensure that date sent and date received and, if applicable, time sent and time received are added as metadata to the associated record profile.*	E		Time references are fundamental to support the evidential value of an e-mail record. It is unthinkable not to appropriately register these metadata elements.
3.2.22	*The RKS must add to the record profile of an e-mail record all the headers of the e-mail associated with that e-mail record.*	E		Any header associated with an e-mail record is a metadata element that must be registered the way it is in order not to weaken the evidential value of the e-mail record. Non compliance risks are therefore self-evident.

3.2.23	*The RKS staff must draw up a list of the file formats that can normally be kept in the RKS. The list must be updated on a regular basis. If an e-mail record includes an attachment in a file format that is not present in the list, the RKS staff must strive to convert the attachment into one of the file formats of the list. The list must differentiate the acceptable file formats according to the requirements established as a result of specific retention and disposition schedules and, if applicable, of other relevant provisions. On-line cross-references from the list to the retention and disposition schedules and any other relevant provision must be available. The RKS must be connected with the list, both to enable the RKS to prompt possible decisions and to make the information available for reference.*	E		Such a tool is essential not only for long-term preservation purposes, but also to support the usability and reliability of every e-mail record throughout the retention period established by the relevant retention and disposition schedule. Non-compliance means that a fundamental resource is not available for the RKS staff.
3.2.24	*During the phase of capture and filing, any e-mail record must be assigned a unique identifier and must be classified, filed, associated with a retention and disposition schedule and handled just as any other record captured in the RKS.*	E		Failure to comply with this recommendation simply means that the integration of the e-mail records in the RKS has not taken place.
3.3.1	*The RKS must keep track of any event affecting an e-mail record included in the RKS. To this purpose, appropriate audit trails must be kept. The audit trail of an e-mail record must be able to indicate which digital components of the e-mail have been affected by the event. An audit trail of each digital component of the e-mail record must be also kept. The record profile of an e-mail record must show the audit trails available for the record and for its digital components. The audit trails must be considered, to all intents and purposes, as a part of the metadata of the e-mail record.*	E		The monitoring of all the actions performed on a digital record is essential. Risks of non-compliance are therefore self-evident.

3.3.2	*Whenever an e-mail record is captured, the RKS must create a checksum of the whole e-mail record and a checksum for each of its components. All these checksums must be stored as metadata.*	E		Checksums are necessary to support the evidence of the integrity of an e-mail record. Non-compliance makes it far more difficult to demonstrate such integrity.
3.3.3	*Whenever even only a component of an e-mail record is modified, the RKS must create a new checksum of the e-mail record and a new checksum for any of its components that has been changed or modified. All these checksums must be stored as metadata in addition to the old checksums.*	E		This recommendation is necessary to properly implement recommendation 3.3.2. Non- compliance risks are therefore the same as those of the former recommendation.
3.3.4	*The RKS personnel must set rules with regard to the use of the forward and reply functions to create outgoing and internal e-mail records.*	E		If these rules are absent, an improper use of thread functions might lead to the creation of e-mail records containing indistinguishable parts of other e-mail records or documents: the negative consequences of such a situation are obvious.
3.3.5	*If the RKS staff notice that an incoming e-mail record contains parts of other e-mail messages difficult or impossible to be correctly identified, the RKS staff must annotate that in the record profile.*	E		An organization and its RKS staff must always do their best to properly describe the conditions under which a given transaction has taken place. Non-compliance occurs if it is impossible to do so.
3.3.6	*Appropriate policies and business process workflows should be set to try to avoid that the incoming e-mails may be created through an improper use of the forward and reply functions.*		A	Hopefully, an organization should be able to enforce all the rules that have been set to properly produce outgoing and internal e-mail records. As for incoming e-mails, the creation of which is out of the control of the organization, a more proactive approach is needed. Non-compliance means that an organization gives up any attempt to exert some influence on the way incoming e-mails are made and consequently to somehow facilitate the correct management of records associated with incoming e-mails.

3.3.7	*Regardless of any association existing with a thread of e-mails, any e-mail record must always be included in the RKS in a way that mirrors its intellectual relationships with all the other records of the RKS and the whole of the records kept in the RKS. All the information concerning its association with the thread to which it belongs must be stored in its metadata.*	E		This recommendation is similar to recommendation 3.2.2, in the sense that it is necessary to keep track of both the intellectual and physical relationships of an e-mail record. Failure to comply with this recommendation automatically makes it impossible to appropriately manage e-mail records and leads to the loss of essential metadata.
3.4.1	*The solutions and procedures aimed to ensure the long-term preservation of e-mail records must be developed within a framework of strategies based on migration and use of standard formats and software- and hardware-independent methods of representation of information.*	E		Non-compliance, implies following different approaches, none of which seem to be as reliable as the ones proposed in the recommendation.
3.4.2	*Regardless of any specific solution adopted to further the long-term preservation of e-mail records, the RKS must preserve all the digital components of e-mail records in formats that allow to efficiently and effectively carry out the routine business activities of the organization.*	E		To develop an action for long-term preservation to the detriment of routine business activities does not make sense at all and, moreover, stirs up unsupportive behaviour and even hostility in a work environment towards the program for long-term preservation.
3.4.3	*The organization's staff should use only formats suitable for long-term preservation to create outgoing or internal e-mail records designed for permanent retention.*		A	Non-compliance might imply inability to cope with serious problems of migration as well as of software and hardware obsolescence, while it should be relatively straightforward for an organization to set constraints with regard to the formats to be used for the creation of outgoing or internal e-mail records. However, the consequences of a failure to comply with this recommendation must be assessed on a case-by-case basis.
3.4.4	*The RKS staff should try to enforce the use of formats suitable for long-term preservation for incoming e-mail records designed for permanent retention, by resorting to policies and business process modelling.*		A	Non-compliance risks are similar to those of recommendation 3.3.6: failure to comply with the recommendation means that an organization gives up any attempt to exert some influence on the way incoming e-mails are made and, consequently, to somehow facilitate the correct management of records associated with incoming e-mails and intended to be selected for long-term preservation.

3.4.5	*When an e-mail record designed for permanent retention is captured into the RKS, the RKS should convert all the digital components of the e-mail record encoded in formats that are not suitable for permanent retention to formats fit for long-term preservation purposes.*		A	If the conversion to long-term preservation formats is delayed until the accession to the digital repository, the RKS staff risk being unable to carry it out properly, since some resources that are essential for the process could be no longer available.
3.4.6	*When an e-mail record designed for permanent retention is captured into the RKS, the RKS should link together all the digital components of the e-mail record, all the metadata describing them and other possible digital components and metadata added to the e-mail record for long-term preservation purposes in an OAIS-compliant AIP.*		A	OAIS-compliant AIPs logically, and possibly physically, link together all the digital components of an e-mail record and all the metadata describing them. OAIS is the conceptual framework chosen by most of the research projects for e-mail records long-term preservation and has been termed by the authors of the Preservation Task Report of InterPARES 1 as "the basis for the content of the preservation process model" (InterPARES 1, "Preservation Task Force Report", 2001, 9. http://www.interpares.org/book/interpares_book_f_part3.pdf - last accessed April 1, 2010). Non-compliance implies following different models, none of which seem to be as reliable as OAIS. Moreover, the sooner OAIS compliant AIPs are created, the better, since some resources that are essential for the process could be no longer available afterwards.
3.4.7	*When an e-mail record that must be permanently retained is moved from the RKS to the digital repository, the RKS must convert all the digital components of the e-mail record encoded in formats that are not suitable for permanent retention to formats fit for long-term preservation purposes. The RKS staff must ensure that the RKS preserves, throughout all the active phase of the life cycle of the e-mail record, all the metadata and the resources necessary to correctly perform the conversion.*	E		This recommendation is really one of last resort. Failure to comply with it makes impossible any action for long-term preservation.

3.4.8	*When an e-mail record designed for permanent retention is moved from the RKS to the digital repository, all the digital components of the e-mail record, all the metadata describing them and other digital components and metadata added to the e-mail record for long-term preservation purposes must be logically linked together in an OAIS-compliant AIP.*	E		Non-compliance means that an organization has decided not to avail itself of the OAIS model. This choice involves risks: see above recommendation 3.4.6.
3.4.9	*The RKS staff must set up an appropriate procedure to ensure the ongoing maintenance of e-mail record OAIS-compliant AIPs and schedule all the operations required to continuously guarantee their usability and trustworthiness.*	E		OAIS-compliant AIPs need regular maintenance: subsequent migrations, inclusions of new metadata, upkeep of audit trails. Failure to abide by this recommendation implies that AIPs are doomed to become outdated and useless.
3.4.10	*All the metadata documenting the custodial history and the maintenance of an e-mail record OAIS-compliant AIP must be added to the AIP.*	E		Failure to comply with this recommendation will seriously impair the trustworthiness of AIPs.
3.5.1	*At the time of the capture of an e-mail record, the RKS staff must appropriately describe every link connected with an external resource. The description must at least include the URI of the resource, the outcome of a check test on the link, the date and time of the check test and the juridical or physical person accountable for the external resource. If necessary, the description must also include any other kind of metadata that turn out to be useful to give evidence of the configuration of such a resource at the time of the capture of the record in the RKS.*	E		All the external resources linked to an e-mail record must be adequately described, otherwise the evidential value of the e-mail record is likely to be seriously impaired.
3.5.2	*The RKS staff must be able to help the organization to specify in which cases it is inappropriate that determined elements essential to use an e-mail record created in the course of a given transaction are kept in an external resource, and must try to prevent such cases by resorting to suitable policies and business process modelling activities.*	E		Non-compliance might result in the creation of e-mail records that are of no use or lacking in evidential value.

3.5.3	*The RKS staff must ensure the validity of any digital signature appended to a digital component of an e-mail record by preserving the original bitstream of the digital component and performing all the actions required to operate the validation chain of the digital signature, as long as the validity of the signature must be retained because of legal obligations and business needs. The RKS staff must register in the record profile the metadata needed to appropriately describe the digital signature.*	E		The negative consequence of non-compliance is self-evident: incapability to meet the legal obligations and business needs that are the reason for the inclusion of a digital signature in an e-mail record.
3.5.4	*If a digital component is associated with a digital signature and the e-mail record to which the digital component belongs has been selected for permanent retention, the RKS staff should convert the digital component to a format fit for long-term preservation and to be kept together with the original bitstreams of the component and the digital signature as soon as the e-mail record has been included in the RKS, and register in the record profile the metadata needed to appropriately describe the traces of the digital signature.*		A	Much as in the case of recommendation 3.4.5, if the collection of metadata and the conversion to long-term preservation formats are delayed until the accession to the digital repository, the RKS staff risks being unable to carry them out properly, since some resources that are essential for the process could be no longer available.
3.5.5	*If a digital component is associated with a digital signature and the e-mail record to which the digital component belongs has been selected for permanent retention, the RKS staff must convert the digital component to a format fit for long-term preservation and to be kept together with the original bitstreams of the component and the digital signature as soon as the e-mail record has been moved from the RKS to the digital repository and the digital signature is no longer needed to meet the legal obligations and business needs by reason of which it has been created, and register in the record profile the metadata needed to appropriately describe the traces of the digital signature.*	E		Failure to comply with it makes impossible any action for the long- term preservation of an e-mail record with which a digital signature has been associated and the collection of metadata elements that may adequately describe the digital signature.

3.5.6	*As long as the digital signature of a digital component of an e-mail record continues to be valid, the RKS staff must register in the record profile the metadata needed to appropriately describe any action and modification performed to operate the validation chain of the signature.*	E		The negative consequence of non-compliance is self-evident: inability to properly demonstrate that the digital signature has been keeping its validity throughout the time it serves to meet the legal obligations and business needs by reason of which it has been created.

BIBLIOGRAPHY

1) Agencies of Washington State Government, Office of the Secretary of State Division of Archives and Records Management. *General Records Retention Schedules*. Olympia, WA: Washington State Records Committee 2005. http://library.evergreen.edu/recordsmanagement/RecordsSchedules/WAGSRet.pdf (last accessed June 2, 2011).

2) The Associated Press. "Sarah Palin e-mail ruling allows use of private accounts to conduct Alaska state business." Nola.com, August 13 2009. http://www.nola.com/news/index.ssf/2009/08/sarah_palin_email_ruling_allow.html (last accessed March 17, 2010).

3) Boudrez, F. "Digital Signatures and Electronic Records." *Archival Science* 7 (2007).

4) Boudrez, F. *Filing and Archiving E-Mail*. Antwerp: Expertisecentrum DAVID vzw 2006.

5) Digital Preservation Testbed. *From digital volatility to digital permanence. Preserving email*. The Hague: Digital Preservation Testbed 2003.

6) Digital Preservation Testbed. *Technical Description TestbedXMaiL*. The Hague: Digital Preservation Testbed 2005.

7) Department of Defense. *DoD 5015.2 STD, Design Criteria Standard for Electronic Records Management Software Applications*. Assistant Secretary of Defense for Networks and Information Integration / Department of Defense Chief Information Officer 2007. http://jitc.fhu.disa.mil/recmgt/p50152stdapr07.pdf (last accessed March 23, 2010).

8) Duranti, L., T. Eastwood and H. MacNeil. "The Preservation of the Integrity of Electronic Records. Rules for Activities Involved in Create records, Handle Records, and Preserve Records." SLAIS- UBC, rule A34. http://www.interpares.org/UBCProject/rules2.htm (last accessed March 30, 2010).

9) Duranti L. and R. Preston, eds. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Rome, Italy: ANAI 2008.

10) Entlich, R. "You've Got Mail—Now What? Regulatory and Policy Dilemmas in Email Management. Part II. US State Environment." *RLG DigiNews*, 2006. http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3019.html and http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3021.pdf (last accessed March 18, 2010).

11) European Commission. "Model Requirements for the Management of Electronic Records (MoReq2)." Project Consult. 2008. http://www.project-consult.net/Files/MoReq2_body_v1_0.pdf (last accessed March 20, 2010).

12) Galloway, P. , M. Green, S. Gunn and S. Soy. "Coming to TERM Designing the Texas Email Repository Model." *D-Lib Magazine* 8 no.9 (September 2002). http://www.dlib.org/dlib/september02/galloway/09galloway.html (last accessed March 29, 2010).

13) Groven, A.-K., J. Ølnes, H. Abie and T. Fretland. *Preservation of Trust in Long-Term Records Management Systems. A State of Art Overview for the LongRec Project*. Oslo: Norwegian Computing Center 2008. http://research.dnv.com/longrec/Intranet/ResearchResults/StateOfTheArt/longrec-trust-soa-report- final.doc (last accessed June 2, 2011).

14) Hornug, M.S. "Think Before You Type: A Look at Email Privacy in the Work Place." *Fordham Journal of Corporate & Financial Law* 11, no. 1 (2005). http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1205&context=jcfl (last accessed May 15, 2011).

15) Huth, G. *Managing E-mail Effectively*. Archives Technical Information Series 62. Albany, NY: New York State Archives 2002. http://www.archives.nysed.gov/a/records/mr_pub62.pdf (last accessed March 23, 2010).

16) InterPARES 1 Project. "Preservation Task Force Report", 2001. http://www.interpares.org/book/interpares_book_f_part3.pdf (last accessed April 1, 2010).

17) InterPARES 3 Project, TEAM Canada. "Template for Diplomatic Analysis." http://www.interpares.org/display_file.cfm?doc=ip3_template_for_diplomatic_analysis.pdf (last accessed March 19, 2010).

18) InterPARES 3 Project, TEAM Canada. "Intellectual Framework Version 2.0." September 2008. http://www.interpares.org/display_file.cfm?doc=ip3_intellectual_framework.pdf (last accessed March 19, 2010).

19) Komanecki, C. "The Move Toward Mandatory Encryption of Sensitive Personal Information." World Law Group. 2. http://www.theworldlawgroup.com/files/file/docs/Faegre%20Encryption.pdf (last accessed May 12, 2011).

20) Lavoie, B.F. "The Open Archival Information System Reference Model: Introductory Guide." DPC Technology Watch Series Report 04-01. OCLC and DPC 2004. http://www.dpconline.org/component/docman/doc_download/91-introduction-to-oais-introduction- to-oais (last accessed June 2, 2011).

21) Paradigm, "Workbook on Digital Private Papers. Exporting email from Microsoft Outlook email clients." http://www.paradigm.ac.uk/workbook/accessioning/microsoft-outlook.html (last accessed March 23, 2010).

22) Pennock, M. *DCC. Digital Curation Manual. Instalment on 'Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages'*. Digital Curation Centre ed. Bath: University of Bath 2006. http://eprints.erpanet.org/113/01/curating-e-mails.pdf (last accessed March 28, 2010).

23) Pennock, M. "Managing and Preserving E-mails", UKOLN, 2006. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/RMS-b_mngmt-pres-emails.pdf (last accessed March 30, 2010).

24) New Zealand Archives. "Fact Sheet E-mail." 2006. http://continuum.archives.govt.nz/files/file/factsheets/f10.pdf (last accessed March 29, 2010).

25) Piccinni, M. L. "Privacy ed e-mail aziendale, una guida per capire." *SearchSecurity.it*, 2009. http://searchsecurity.techtarget.it/articoli/0,1254,18_ART_104167,00.html?lw=18 (last accessed March 17, 2010).

26) Plouin, G. "L'e-mail, condamné à évoluer ou à disparaître," *Le Journal du Net*. http://www.journaldunet.com/solutions/0506/050610_tribune.shtml (last accessed March 19, 2010).

27) Pontevolpe, G. and S. Salza. *General Study 05 – Keeping and Preserving*. Version 4.0. The InterPARES 3 Project, Team Italy: June 2009. http://www.interpares.org/rws/display_file.cfm?doc=ip3_italy_gs05_draft_report_v4_RESTRICTE D.pdf . Restricted draft, i.e., access to the file is restricted to the InterPARES participants (last accessed June 2, 2011).

28) Potter M. "Researching Long Term Digital Preservation Approaches in the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring)." *RLGDigiNews* 6, no.3 (June 15 2002). http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2974.html (last accessed March 31, 2010).

29) Privacy International. "PHR2006 - Privacy Topics - Workplace Privacy#Email, Internet Use and Blog Monitoring." 2007. http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-559090 (last accessed March 17, 2010).

30) Queensland State Archives. "Managing Emails that are Public Records." 2007. http://www.archives.qld.gov.au/downloads/emails_that_are_public_records_policy_and_guideline.p df (last accessed March 23, 2010).

31) Radicati Group Inc. *Microsoft Exchange Market Share Statistics, 2005*. Palo Alto, CA: The Radicati Group July 2005. download.microsoft.com/download/E/8/A/E8A154BF-CC35-4340- BD26-6265CDB06B6E/ExStats.doc (last accessed March 16, 2010).

32) Roberts, A. *Blacked Out. Government Secrecy in the Information Age*. Cambridge, UK: Cambridge University Press 2006.

33) Rosencrance, L. "Sidebar: Employee Rights and Relations." Computerworld, 2004. http://www.computerworld.com/s/article/93472/Sidebar_Employee_Rights_and_Relations?taxono myId=018 (last accessed March 17, 2010).

34) Santangelo, J. "Rise of the Machines: The Role of Text Analytics in Record Classification and Disposition." *Information Management Journal* 43, no. 6 (November-December 2009): 22-26. http://content.arma.org/IMM/Libraries/Nov-Dec_2009_PDFs/IMM_1109_rise_of_the_machines.sflb.ashx (last accessed March 26, 2010).

35) The Sedona Conference. *The Sedona Canada Principles. Addressing Electronic Discovery*. The Sedona Conference 2008.

36) State of Wisconsin, Email Policy Task Force. "Literature Review- Professional standards, academic best practices." Department of Administration of the State of Wisconsin. www.doa.state.wi.us/docs_view2.asp?docid=6079 (last accessed March 23, 2010).

37) State Records Authority of the New South Wales. "Government Recordkeeping Manual. Rules. Polices. Policy on Electronic Messages as Records." 1998. #3.1. http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/rules/policies/policy-on-electronic-messages-as-records (last accessed March 23, 2010).

38) State Records Authority of New South Wales. "Government Recordkeeping Manual. Guidance. Guidelines. Guideline 13 - Create and capture: Guidelines on better recordkeeping", http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping- manual/guidance/guidelines/guideline-13. 2001. #4.2 and #4.3 (last accessed May 4, 2011).

39) Texas Digital Library. "Vireo Email Templates." http://wikis.tdl.org/tdl/Vireo_Email_Templates. 2010. (last accessed May 4, 2011).