ARCHIVI & COMPUTER
AUTOMAZIONE E BENI CULTURALI

Anno XIV — Fascicolo 3/04

Dai testi unici ai codici: gli archivi e la nuova normativa

INDICE

SAGGI

Maria GUERCIO	*Dai testi unici ai codici: gli archivi e la nuova normativa*	p. 7
Paola CARUCCI	*La protezione dei dati personali, l'accesso ai documenti amministrativi e la consultabilità degli archivi storici*	10
Maria Grazia PASTURA	*Tra codice dei beni culturali e codice della privacy: cosa cambia nella disciplina di tutela, conservazione e valorizzazione degli archivi e nel diritto di consultazione e di accesso*	37
Guglielmo LONGOBARDI	*Le tecnologie dell'informazione e della comunicazione come strumento di democrazia: l'accessibilità*	49
Francesco GRASSO Guglielmo LONGOBARDI	*Il software open source*	55
Maurizio VITTORIA	*La nuova normativa in materia di accessibilità del web. Una valutazione concreta*	66

INTERVENTI

Michele DI SIVO	*Archivisti d'inizio secolo: antichi strumenti, nuovi linguaggi. Normalizzazione e condivisione degli inventari: la proposta dell'Archivio di Stato di Roma*	74

"eccessiva" e frenare la ricerca di ciò che non risulta dalle loro descrizioni. Questo tipo di condizionamento è davvero rischioso in quanto può fuorviare lo studioso, diseducarlo e portarlo a ritenere che quanto espresso dall'inventario e dalla banca dati esaurisca il contenuto dell'archivio. Occorre insomma saper comunicare anche l'incompletezza: dai documenti possono giungere risposte che nessun inventario e nessun indice – per quanto ben fatti – sono in grado di dare. Il disorientamento non è legato alla forzata uniformazione dei vecchi inventari – un sistema informativo si costruisce anche per integrare dati difformi –, ma al rischio che il mezzo informatico tagli fuori la figura dell'archivista-mediatore.

Per quanto si possa fornire il sistema di testi che spieghino il valore dei mezzi di corredo e le procedure della ricerca, nessuna applicazione informatica può sostituire l'intervento dell'archivista che discute con il ricercatore e regola la sua risposta in base al tipo di domanda. Nel caso dell'accesso remoto la cosa si complica, ma il ricercatore dovrà comunque avere la possibilità di usufruire della mediazione; ciò che cambierà sarà (e in parte già è così) l'ambiente di tale mediazione: l'e-mail che sostituisce o si integra con la telefonata o con il colloquio in sala di studio, lo scambio di testi o di ipertesti nel web non solo divengono parte importante del nostro lavoro, ma ne modificano progressivamente l'organizzazione, a cominciare dalle questioni relative all'archiviazione dei documenti informatici prodotti attraverso tali scambi.

Il ruolo dell'archivista diviene dunque più necessario per molte ragioni nuove. Deve affermare il principio di provenienza in ambiente informatico, dove è forte la tendenza alla decontestualizzazione. Deve utilizzare strumenti e linguaggi affini a quelli della catalogazione bibliografica, ma mantenere l'identità della cultura archivistica. Deve sostenere una domanda di ricerca che cresce e che cambia di qualità, si allarga a soggetti non specialisti e all'accesso remoto. Dovrebbe infine trasferire le conoscenze acquisite alle generazioni successive. Deve insomma affrontare un nuovo "sbandamento di idee": anch'esso, come alla fine del Settecento, apre un lungo e accidentato percorso e come allora è la diretta conseguenza di un'estensione della pubblicità degli archivi. I suoi esiti non sono del tutto prevedibili, ma la sfida è irrinunciabile.

Hans HOFMAN

Can bits and bytes be authentic? Preserving the authenticity of digital objects[1]

Introduction

At a round table, organised by DigiCULT in May 2002 in Barcelona, representatives of different communities discussed the issue of authenticity of digital information. It turned out to be not easy to understand each other, because of different terminology, or better the use of the same words, but with different meanings, a familiar problem in discussions across domains. One thing was clear, however: it is a relevant issue not limited to one community, but shared by many, as for instance the libraries, broadcasting, and archives communities. Authenticity with respect to documents or information entities is not a new notion. It goes back as long as they exist. In a paper world with physical entities it was not a big issue, because of the fixed form they had. The discussion however, showed that the emergence of information in digital form requires re-thinking in applying it in the digital world. Authenticity may be described as 'conveying a (digital) object in its 'original state' to a user over time'. Although this may not be an adequate description, it already implies a lot of issues, some of which I want to address in this paper. One of them concerns the issue of what 'original' means in a digital environment. One of the recommendations of the round table was to conduct a survey to see what is happening in practice and how authenticity is experienced or seen in different communities, especially in relation to digital objects. It also indicates that no consensus has been reached yet on this issue[2].

One conclusion can be drawn, which is that authenticity is an important requirement for preserving information, especially in a digital environment. Research on

[1] Revised paper of a presentation at the annual IFLA conference in Glasgow, 23 August 2002. Hans Hofman is co-director of the European project ERPANET and working as a senior advisor at the National Archives of the Netherlands and involved in several (national and international) projects in the area of digital preservation, such as the InterPares project and the ISO TC46/SC11 records management committee, in particular as chair of the Working Group on records management metadata.

[2] For a report on this round table, see www.digicult.info and the DigiCULT publication *Integrity and authenticity of digital cultural heritage objects*, August 2002, (Thematic Issue 1).

digital preservation has been dominated so far by libraries, certainly if one takes into account what is published in this respect. A central position in current research is taken by the Open Archival Information System (OAIS), a reference model for digital preservation, which is now an ISO standard. Again the main communities involved in this effort are the library and scientific communities, with only minor representation from archives. In order to get a more balanced and broader view more insight in archival approaches may therefore be useful.

In this paper I will try to discuss some of the issues involved, such as what does authenticity mean in a digital world, and what we want to preserve, if we want to preserve the 'original object'. The other main issue is, how do we ensure authenticity for digital objects, and finally, will answers to these questions differ for different communities?

Authenticity and digital preservation

Digital preservation can be described in different ways. One of them is 'ensuring the authenticity, usability, and accessibility of digital objects through time as long as required'. Another one is 'ensuring technological survival of digital information or objects as long as required'. Both may be valid. The first one is a more broad definition that is taking into account more the intellectual aspects, the second one is narrower and looks more at the technical aspects. Both views are necessary and even complementary. They identify the aspects or requirements we have to consider when preserving digital objects. Here I will discuss two key issues: authenticity and digital objects.

Authenticity is said to be a big issue in preserving digital information. In a physical paper world it is too, but there it is not really an issue. Paper is fixed and its very physical features already provide us with (relevant) information to assess whether it is authentic or not. Nonetheless forgery is of all ages, and as we know also present in a physical world, so there are methods developed to identify the authenticity and integrity of paper or other objects.

From paper to digital we are entering a new world that does not have established practices and rules yet with respect to identifying authenticity. Contrary to the physical paper world digital objects can be altered relatively easy. One consequence is that a new infrastructure (e.g. technical, organisational) has to be established and implemented to prevent that. The different nature of digital objects requires furthermore a translation of the existing concept of authenticity, so it can be applied to digital entities.

The issue of authenticity and integrity

Authenticity is often described as 'an object is what it purports to be'. The basic notion of authenticity is that we trust things and persons as they present themselves to us as long as the opposite is not proved. Authenticity can also be described as 'to

be true in its own context'. Meaning that if we know when, where and why something is created, we are able to assess its authenticity.

The question is, can we in a digital world? Let me give an example. Recently I searched for a page of a (Dutch) website that has disappeared since 2 years now, by using the so-called Wayback Machine[3]. After some search I discovered and retrieved several occurrences of that website (or better webpages) that were captured during a 3-year period. At first sight the retrieved website looks similar to what it once was, that is what I remembered of it. At second sight and after some study it was not the same website, however. It seems as if you access one website, but you don't. That can be derived from the date on which the Homepage was harvested. That turns out to be different from the date of the underlying pages. They are harvested on different dates and that goes for each 'layer' of pages of the website. So in fact I navigate through the presented 'website' that consists of pages from different dates (which are at random). That means it does not reflect the situation of the whole website at one point in time. When I retrieve and access the webpages, however, I do not know that and, worse I am not even getting that information. Only if I have a look at the URLs that are presented I can discover the different dates. Although the 'internet archive' as presented by the Wayback Machine does not explicitly state that the webpages showed are authentic, it is the implicit assumption of it that makes it an issue. The Homepage and underlying pages are thus not a coherent representation, but a reflection over a random period of time. What is needed is explicit information about the provenance and structure of the webpages presented to the user, which will enable him or her to assess their usefulness and value. It will prevent the misinterpretation of the information.

The fact that the website as presented is a collage and not one coherent information entity does not have to be a problem, if only I know. In the case of the Wayback Machine however I do not get that information (metadata), so I may think it is all one consistent set of webpages (or one website).

The only metadata I get, be it in a kind of hidden way, is the date (in the URL) on which each webpage was harvested. That is not sufficient. There are more shortcomings, inherent to the harvesting technology and that is the fact that in the earliest captured homepage not all pictures were captured (for reasons of efficiency and time) as well as not all external links. The consequence is that I may see webpages in which pictures or logos are missing, and therefore incomplete. Whether this is always relevant for understanding or using the content as offered, is as said another question, but at least information about these inadequacies should be provided.

One criterion for authenticity will be that we should be able to establish the identity of an object, which includes information about the origin or provenance, when and why it was created and used. That is particularly relevant for (archival)

[3] See www.archive.org.

records, which are created in business processes (context information). Records contain information about what happened, but can also serve as evidence of that. If we know where it comes from, we can understand it better and use it more effectively. This identity can also be derived from relationships to other documents or records ('documentary context'). That information will help us in establishing whether we can trust it or not.

The easiness with which information or documents can be altered in a digital environment raises questions about authenticity and about how to deal with that. It is clear that it will have implications for preservation, or, putting it the other way around, it poses one of the main requirements for it.

If we know that an information entity is authentic then the information it contains may be trustworthy. It does not mean automatically that the information is true, however. We may assume that a record is a true representation of the business activity (or transaction) that created it, but there does not have to be a causal relationship. A forgery for instance can also be authentic. It is a forgery, because there is a mismatch between its content and context ('what it purports to be') on the one hand and the creator and the reason why he created it, on the other. It is authentic, because it represents a forgery (it is evidence of forgery).

That poses the question when do we need evidence of authenticity? The answer has been given already, information needs to be trustworthy if we want to use it. This trustworthiness depends on the extent to which we are able to establish where it comes from and assess that it is it says it is. The use can be for a variety of things, such as doing business, control, audit, accountability, re-use etc.

It is also a matter of risk management to identify how serious this authenticity question is and subsequently our authenticity requirements are. The more crucial something is, e.g. in parliamentary inquiries or court cases with high financial stakes, the higher the requirements will be for instance.

It is for each organisation to decide what its requirements for authenticity are in each business activity or of the organisation as a whole. For publications it may be different, in the sense that people want to be sure it is the thing that is once written by a certain person. For computer games it is often a kind of nostalgia or the wish to know how people played them some decades ago.

What do we want to preserve?

The second issue is that should be discussed in this respect is the question what is it we want to preserve, or what is the object we are talking about in a digital environment? Terms like digital object, data object, information object, or information resource are often used, but what do they mean? In digital form there is no physical entity any more. The data files with encoded data as stored on a medium should be saved, but are they the objects we are talking about and is that what we want to preserve? There is a growing awareness that an object has both physical and intellectual aspects, that are equally important to preserve, but also have both an impact on the method of preservation. Let me try to explain this.

In order to present a digital document we need a data file stored on a medium and some software application, running on a hardware platform that is able to interpret the (encoded) bits and bytes and to reproduce the document. The document thus shown on the computer screen is the result of an interaction of these technical components or a process if you like. In this process we can identify the following components: the data file, the medium, the hard- and software and something non-physical or virtual, the document on the screen. It is obvious that the document we view is not similar to what is stored on the medium. In many cases more than one data file even may be necessary to recreate the document. An example is a multimedia document with text-elements and pictures, or text files that need a dll-file (dynamic link libraries, used by the application) to present the proper type fonts in a document. It is also clear that we need all these components to reproduce the (virtual) document. The question is now, what is the digital object here? And if we look at authenticity, what is the object we are talking about in this respect? Is that the data file or will that be the (virtual) document? The terminology is confusing and requires explicit explanation and a clear distinction between technical and intellectual aspects. In many discussions or articles these questions are mostly not addressed, but they are essential for understanding the implications for preservation. To provide an answer: if we talk about authenticity it is the virtual document (in the broadest sense, i.e. including views of a database for instance) as presented on the screen that should be authentic.

Confusion sometimes also exists on what a document still is in a digital environment. Another issue that may arise is how static or fixed are these properties? It seems as if their boundaries are blurring in this new infinite world of information, because of endless links, especially on the web or in a distributed database environment. There is the emergence of new types of documents that have an increasingly dynamic character, as for instance compound documents with hyperlinks in it that may change over time. How do we deal with this fluidity, which can be a relevant aspect of a (web) document? In printed form the hyperlinks in the document will be fixed, but the information about the link, as to what other document or web page it refers, is lost as well, unless some description of the link has been captured. Will that be sufficient or do we still need to have the links 'alive', but then to what extent. On the web the linkage between documents may be endless and where do we stop?

This raises the question as whether we are loosing the concept of a document. Can they still be identified as discrete entities? It may be that people are deluded by the systems in which these documents are created, or in which they are stored as digital objects, and through which they are presented on the screen. Sure, the new digital environment provides us with many ambiguities and all sorts of difficulties

in identifying documents or records, but we have to keep in mind the basic principle of what a record is (in the archival sense). Documents and records are created with an intention and in a certain context. For records that is a business activity (of companies, or government organisations, or persons) and in that context they are confined by the role they play in that business activity and their characteristics are determined by the message they carry or have to convey. These characteristics are implicit and sometimes explicit in the content, form, structure and behaviour they have been given. Documentary forms, such as letters, minutes, resolutions, charts, financial accounts etc., have been developed long ago to help understanding the message and anybody familiar with a business activity and the inherent community conventions knows what they mean. These forms and conventions were simultaneously based upon and bound by the characteristics of paper. Information technology offers other possibilities, and therefore new forms of documents will emerge, e.g. compound, hyperlinked and/or multimedia documents, which make them no longer two-, but three and even multidimensional. It eases the re-use or re-purposing of documents, created in one context and used in many others, adding new layers of meaning and context. It also enables other, new arrangements of documents, or records and information, so new and other user needs can be accommodated. These new dimensions of intellectual relationships force records managers, archivists or other information specialists to broaden their view, to understand the differences with traditional records, and to develop new approaches to deal with them.

So digital objects have both intellectual - the record or publication viewed on the screen - and technical - the digital components (data files and software) that contain the necessary digital representation of it - aspects. The consequence of this notion is that (archival) records and publications are not preserved as such and that preservation of them means to maintain the ability to reproduce them. It also means that there is no longer an 'original record or publication', only a notion of a record or publication as originally created or received and used.

The record seen from an intellectual perspective consists of five main elements as already mentioned: content, structure, form, context and in some cases behaviour. Each of these elements has to be identified and described through metadata, which will explain what should be reproduced (the essential characteristics or properties), the interrelationships these elements and the digital components that contain the encoded data, their provenance, and finally their interrelationships with other related intellectual objects. This last requirement refers to the interrelationship of records as created and received in carrying out a business activity. The records for instance in a case file contain not only information about a specific case individually, but also and more important as a cluster, which is called the documentary context or the 'archival bond'.

The big issue is however, the question what are the essential characteristics? In both the CEDARS and CAMiLEON projects the term 'significant properties'

is used. It indicates the elements of the object that should be preserved. Different categories are identified in these projects, such as types of metadata, types of format, types of functions, lay-out features etc., distinguishing different levels of abstraction. They refer to both technical and intellectual aspects. Since what we want to preserve is the intellectual content with its relevant structure and form and/or behaviour, the focus should be on describing them, and subsequently on the technical implications. Describing these significant properties or essential characteristics requires a kind of formal language, so they can be articulated in a consistent way.

The fact that we are establishing the essential characteristics of digital documents, may sound like we are dealing with a kind of appraisal decision at the level of elements in a document is being made. On the other hand in identifying what kind of documents or records should be created in a certain business process or in our choice for publishing a special type of book or publication we already have made decisions in this respect. The difference is the moment in time we are doing it. The general idea is that especially in a digital environment these decisions should be made beforehand, because they also may or will influence the functionality of software applications.

Roles and responsibilities

This entails also the issue of who will determine what the essential characteristics are? And that brings us to the question, who are involved? Different roles can be distinguished around creating, managing and preserving digital objects. First of all there is the creator who has a certain intention in creating a document or publication through which he or she wants to communicate a 'message' to other people. The creative work will be done in a certain context, such as a business activity or an act of creativity (e.g. writing or making). In the first case records will be created, in the latter a book or publication or a work of art. The second role that can be distinguished is that of the user. He or she may receive or retrieve a digital object within the same timeframe or within the same domain (e.g. a business activity), but it may be also in the future and/or in another domain. That mostly means that it will be used and interpreted from another perspective and in another context, e.g. for research or accountability. The more time or distance exist between creation and use, the more difficult it may be to understand the original intention and context. In the case of medieval documents for instance one will need some knowledge about that time and its conventions and will need to learn the different semantics of words and even to read the handwriting. The point is that in order to use an information source effectively it is necessary to know where it comes from, to be able to trust and to interpret it. The third role we can identify is that of the preserver who is responsible for maintaining the object in its 'original' state over time. The preserver is the intermediate between the creator and the user, if there is a distance across domains and

in time. That means he or she does not only have the responsibility to maintain the information object, but also to keep it meaningful. One organisation or person may fulfil the above mentioned roles at the same time, but as soon as time progresses different people or organisations will be involved. The preserver for instance may be another organisation and the user will be from another domain.

The following diagram shows it in schematic form.

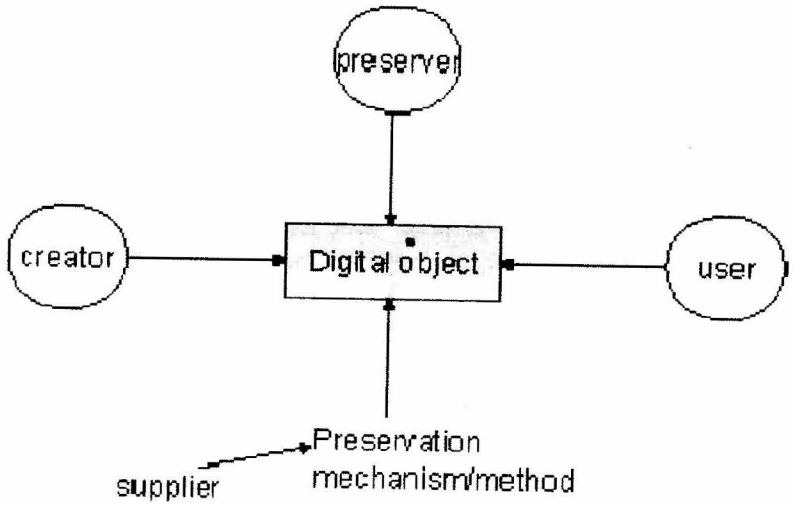

Figure 1 Roles around a digital object

Each of these different roles may or will have a different perspective on the digital object, since there are different responsibilities, but also the context in which a role is played will differ. Does that mean that authenticity will be seen or experienced differently? The result of the act of a creator will be by definition authentic. It is the 'original' object and it is created for or with a certain purpose. The user however may have a slightly different view, because he or she may use it in a different context and time and then the emphasis may lie on other aspects or properties of the object than in the case of the creator. Nonetheless the user wants to be sure that what is shown to her/him is the 'original' object and that it is what it purports to be. He also has to know why it was created in order to assess its value and not to misinterpret the content. So in fact both viewpoints are valid, but only in their own context, i.e. the context in which they are used.

The task of the preserver is to manage and preserve the digital objects in an authentic, usable and accessible way. The above described notion of authenticity will determine what the preserver has to do in order to preserve the digital components and to be able to provide (or reproduce) authentic, reliable and accessible representations of the intellectual objects. In this respect he needs a preservation mechanism, i.e. a coherent set of technologies, facilities, people, systems, processes and procedures. Here we may distinguish even a fourth role, that of the supplier (e.g. a soft- or hardware vendor), that is relevant for implementing the right infrastructure. Software suppliers can provide them if they know what the requirements are. At the moment the multiple available applications and the fact that many of them are proprietary, however, are big issues to be dealt with in relation to authenticity and preserving it.

An essential part of this 'infrastructure' should be a comprehensive preservation policy, consisting of frameworks, strategies and methods for preserving digital objects. Such a policy will be built upon requirements, coming either from legislation, the preserving institution, and issues like authenticity and use, and has to encompass all aspects and perspectives as described before. With respect to the digital object and each of the identified roles for instance an organisation has to formulate a policy. Those policies should be based on common principles and together be coherent and consistent, not only if they exist within one organisation, but also and if possible, if more organisations are involved. So if we consider the creator, there need to be for instance policies for creation, collection management, appraisal, capturing or harvesting the digital objects, transfer including identification of authenticity requirements. In the case of the user policies and mechanisms for access to information sources have to be articulated, for rights management and for description and resource discovery. Digital objects require preservation strategies and methods, e.g. migration or emulation or other methods or technologies, but also creation policies or guidelines (e.g. use of open standards, templates for metadata). Preservers have to deal with implementing the requirements, information security, sustainability, metadata management, facility management, etc. Towards suppliers that may mean that standardisation policies will be in place. This whole framework of policies and strategies needs to be maintained, reviewed, and updated continually. That leads us finally to identify a fifth role, that of the auditor or controller who is responsible for quality assurance or compliance to internal and/or external requirements.

Developing policies and strategies

Preservation does not stop at identification and description of processes and/or metadata, and/or requirements. It is only the beginning. For adequate preservation of digital objects a policy framework is needed, that should rule, guide and direct the activities in this area. Such a framework or policy should consist of some basic

components, including responsibilities, organisational structures, people and expertise, technical infrastructure, strategies, rules and procedures.

Starting from the digital object (bottom-up) I will try to build a picture of the components of such a preservation policy. As identified a digital object has both intellectual (or conceptual) and physical aspects, that are closely related. They need to be managed together and coherently in order to reproduce authentic, usable, accessible and understandable information resources. They are ruled by authenticity requirements, as shown in diagram 2.

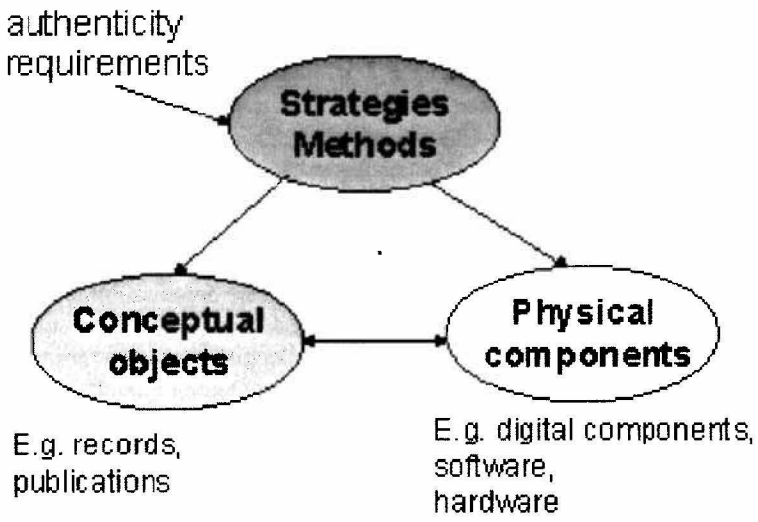

Figure 2 Strategies for objects with conceptual and physical aspects

Since it will be complicated and require much effort to implement comprehensive policies, it may be useful to see what frameworks or models or parts of them are available at the moment, which would enable organisations to build them. They range from high level to very detailed level and each of them on its own will not be sufficient, but together they cover almost the whole area. For the records management community recently the ISO records management standard (ISO 15489) has been published that provides a very useful and comprehensive framework, be it on a high level. For the library community a comparable attempt has been made with the recent OCLC/RLG publication of a report on 'trusted repositories'.[4] In both reports an (pro-) active attitude is recommended, by stating that life cycle management of

digital objects starts with the creation and in the case of (archival) records even before with the design of record creating and management systems. The report on trusted repositories takes the Open Archival Information System (OAIS) reference model as a starting point and intends to be compliant with that model. The OAIS model is an emerging international standard and provides a common framework of processes, terms and concepts, that could help organisations to understand what is needed and in identifying requirements for both people and systems.[5] Six areas for management are identified and described in the RLG-OCLC report:

- *administrative responsibility*, which refers to the commitment to implement agreed standards and best practices, with risk management and contingency planning, and being transparent, and accountable;
- *organisational viability*, which refers to the idea that organisations have the appropriate legal status, staffing and expertise, and review procedures;
- *financial sustainability*, refers to standard accounting procedures and financial planning;
- *technological suitability*, refers to having made open and well-founded decisions, to having in place appropriate hard- and software through time, including policies for replacing it;
- *procedural accountability*, refers to procedures documenting repository practices, monitoring and feedback mechanisms, and
- *system security*, refers to policies for having security procedures and measures in place as well as disaster preparedness.

The idea is that based on the described aspects certification procedures can be developed at different levels. The report suggests in this respect four areas, i.e. individual, institutional, process, and data or reliability certification.

Functionality for preservation systems is described at high level in reference models, such as the OAIS model. For the archival community the preservation function model as developed by the InterPares project may serve as an example[6]. This model is an application of the OAIS model for preserving electronic archival records, except that it goes a little further and does take into account the different aspects of intellectual and physical components. That is not done (yet) in the OAIS model, where digital objects simply are defined as 'objects composed of a set of bit sequences', which is a rather technical approach and limited concept.

[4] RLG/OCLC published a report titled 'Trusted Digital Repositories: Attributes and Responsibilities', May 2002. Available at www.rlg.org/longterm/repositories.pdf.

[5] The OAIS model is mainly addressing processes and their interrelationships and related information entities on a high level, and is not intended to be an implementation model (so it does not contain any requirements at a technical or system design level). See http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf ('blue book').

[6] See www.interpares.org for the final reports on the first Inter Pares project (1999-2001).

The description of processes necessary for preservation of digital objects and how they are related helps organisations to identify what is needed and to formulate strategies and methods. Standardisation, e.g. for storage formats or metadata structures, may be part of these strategies. In the area of preservation methods conversion, migration, and emulation are still seen as the main approaches for preserving digital objects. They do not exclude, but may be complementing each other. There is in this respect a growing consensus that a suite of such methods may be necessary to accomplish the preservation of digital objects. Although some people think that the solution for digital preservation already has been found in the Universal Virtual Computer (UVC) concept, this still has to be proven[7].

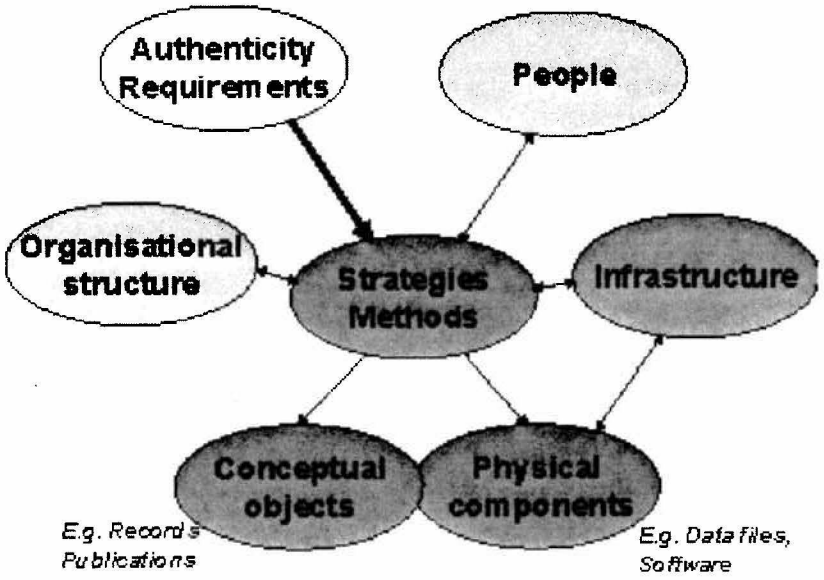

Figure 3. Areas and/or objects that should be covered by a preservation policy

[7] See H.M.G. Gladney in the "Digital Document Quarterly (DDQ)": http://home.pacbell.net/hgladney/. Both the Dutch Royal Library and the Dutch Testbed project are preparing or conducting experiments with this approach. See www.kb.nl/kb/resources/frameset_kenniscentrum-en.nl en www.digitaleduurzaamheid.nl.

In formulating policies and strategies risk is one of the criteria. The more important some piece of information is, the more we depend on authenticity and adequate systems and procedures for preserving it. It requires risk management, as is also indicated as an issue in the above mentioned OCLC/RLG report on 'Trusted Repositories'.

A crucial component of policies is staff, because in the end they are responsible and can make things a success or a failure. That will depend on their awareness of the issues at stake, starting with managers and their commitment to deal with them, but also with professionals and scientists and their available expertise, and the proper responsibility that is assigned to them. It also requires recruitment of the right people, adequate training and last but no least expertise outside traditional disciplines. It will not be sufficient to have librarian or records management expertise, but requires also knowledge of IT and its implications and consequences for managing the digital objects to be preserved. This interdisciplinary co-operation is essential for successful preservation.

So in all policies have to address a whole range of areas that are closely connected and interrelated, as is shown in figure 3. Policies will furthermore be governed by legislative or regulatory frameworks. Sound and robust policies require among others a clear concept of authenticity requirements and of the nature of digital objects or entities. Only then it is possible to train people, to implement an adequate infrastructure and to identify the required strategies and methods for preservation.

In the InterPares project an attempt has been made to identify the authenticity requirements for electronic records and the implications of them are included in a preservation function model. In that model also both the intellectual and physical aspects are included. The OCLC/RLG 'Trusted Repositories' report lacks such an authenticity concept or requirements, and can only provide a high level framework, that still has to be further refined if applied by an organisation responsible for preserving digital objects.

Summary

Preservation of digital objects has to do with the notion of authenticity as one of the major requirements. It seems however that there is not yet a common understanding of this notion applied to these objects. Different communities may have different interpretations, which may lead to different solutions. In general however a common notion exist, that has to be expressed in a language that is understandable for all communities involved, and in that respect more effort and progress are needed.

In summarising the above a few things can be said. Understanding authenticity of digital entities requires understanding of the nature of digital objects. These objects have both intellectual and physical aspects. In the many existing publications

this distinction is mostly not made, unfortunately because it is an essential notion for preserving 'digital objects', whatever they may be (records, or publications, or other kinds of information resources). Authenticity has to do with the intellectual aspects of these objects. This concept of authenticity is also not 'static' in the sense that it will be similar to everybody. Depending on the role (and perspective) there may be different essentials. The view of a creator for instance may differ from that of a user.

One of the concepts that is lost in a digital environment, is the notion of the 'original'. Preservation of digital objects means maintaining the ability of reproducing these objects and especially the intellectual aspects ('construct') of them. The concept of authenticity therefore needs a new implementation in the virtual digital world.

In preserving digital objects a coherent and overall policy, including strategies, procedures, and methods, is necessary. This policy encompasses a broad spectrum of different areas, which are closely interrelated. It clearly is also more than identifying a technological solution. New and adequate policies, strategies, and methods that can deal with this new situation and the inherent requirements have still to be developed. Frameworks such as the OAIS reference model and the report on trusted repositories are very useful starting points. They provide no clear-cut policies or solutions, but only a basis for further thinking and application in specific domains. It depends on the authenticity requirements in connection to legislative and regulatory frameworks in those specific domains how policies and strategies should be designed and developed. Nonetheless more discussion between disciplines is needed to get a better view and notion of authenticity and the implications for developing policies and systems. In this respect cross-domain collaboration and fertilisation may help identify common interests, concepts and moreover help to develop the preservation policies and strategies so needed. Collaborative cross-domain initiatives in the area of preservation of digital resources, such as ERPANET can foster this.

Seamus ROSS, Monica GREENAN, Peter MCKINNEY[1]

Strategie per la conservazione digitale: Descrizione e risultati dei primi studi di casi di ERPANET *

(traduzione italiana di Francesca Marini)

Abstract: *ERPANET, a key European Commission Fifth Framework digital preservation activity, is conducting, among its activities, a series of case studies to improve our knowledge about digital preservation practices within Europe's public institutions and private sector companies. The research group has conducted interviews with staff from many companies spread across the broadcasting, pharmaceuticals, publishing, and telecommunication sectors. This paper describes methods, presents the results of the research, and defines areas were the authors believe further work (e.g. research, guidelines, improved practices) is urgently needed.*

Introduzione

Archivisti e bibliotecari sono uniti nel richiedere interventi urgenti che assicurino la sopravvivenza dell'informazione digitale. Per capire in quale modo le organizzazioni europee si stiano occupando di conservazione digitale, ERPANET (Electronic Resource Preservation and Access Network[2]) ha elaborato una serie di studi di casi per indagare sul grado di consapevolezza presente negli enti e nelle

* Questa relazione è stata presentata, in una versione preliminare, al *Preservation of electronic records: new knowledge and decision-making. Symposium 2003*, Ottawa, Canada (Canadian Conservation Institute, 15-18 settembre 2003).

[1] Gli autori fanno parte del Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, www.hatii.arts.gla.ac.uk. Si ringraziano i colleghi di ERPANET Andreas Aschenbrenner (Nationaal Archief Nederland, Den Haag), Georg Büchler (Schweizerisches Bundesarchiv), e Prisca Giordani e Vincenzo Di Meo (ISTBAL, Università di Urbino) che hanno contribuito alla conduzione degli studi di casi.

[2] ERPANET è un'attività finanziata dalla Commissione europea (IST-2001-32706). Si veda www.erpanet.org per ulteriori informazioni e per i materiali a disposizione. I direttori di ERPANET sono: Niklaus Bütikofer (Schweizerisches Bundesarchiv), Maria Guercio (ISTBAL, Università di Urbino), Hans Hofman (Nationaal Archief Nederland, Den Haag), Seamus Ross (HATII, University of Glasgow).