НАУЧНИ РАД

УДК 005.923.2:004]:02
УДК 026:004.738.52

RETHINKING HOW TO SHAPE A NEW MATRIX FOR THE PROTECTION AND RETENTION OF CULTURAL HERITAGE*

Paolo Budroni**
Vienna University, Library and archive services, Project Phaidra

Abstract: Institutional repositories are currently evolving into a single core of digital collections. Necessary for this evolution is the task of charting the principles used in creating a matrix, which can be derived from the rules for the preservation of cultural heritage in order to endure over the course of time. These principles should be derived in such a way that they develop far beyond what the usual technology-dependent advice and methods would. Technologies come and go, but preservation is a task that should be carried out independently of the currently common line of thinking. Technology with know-how in preservation matters is only one aspect. Preservation is a cultural mission, which must guide our actions. The code that we select in the task of preservation, makes our culture bound behavior reproducible.

Keywords: Open Access; open university; information storage; dissemination of information; cultural policy

*This article has been published in Austria in: Mitteilungen der VÖB (Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare) 63 (2010) Nr. 1/2; "Manifest zur Bildung einer Matrix [code], Sicherung und Verschleiß des Cultural Heritage in Europa

**paolo.budroni@univie.ac.at

1. Introduction

In 1984, Italo Calvino was officially invited by Harvard University in Cambridge, Massachusetts, to present the "Norton Lectures" which appeared in the preface of his posthumously published book[1]. These were a series of six lectures focussing on the topic of "Poetry". This referred to any form of poetic communication, namely, of any: *literary*, *visual*, *musical* form. This brings us to the heart of the matter, since the objects of his studies also refer to objects with a multimedia format that are stored in our current digital repositories for posterity. This is especially true for institutional repositories in the scientific establishment, which is now evolving toward a core of digital collections and whose job is the careful storage of cultural heritage, including all teaching and research output in such a way that it remains accessible, usable and understandable over the long term.

Italo Calvino, as is further stated in the preface, was almost obsessed with clarifying this topic. In this passage, he presented some literary values that needed to be preserved for the next millennium. Calvino named these values as: "Six memos for the next millennium." These are in order: lightness, quickness, exactitude, visibility, multiplicity and, not carried out by him, *consistency*. With that which Calvino meant in his Norton Lectures on poetry, the author of these lines identifies himself and thus, makes the following issues the focus of his essay:

What principle is recognized in the field of *preservation of cultural heritage* and more specifically, how can one use it to create a code that is both derived from universal rules and can be reproduced at any time?

What are the key ideas and possibilities available to us that will withstand deterioration?

Technologies come and go. What would be beyond the usual technology-dependent advice and methods of use that would enable us to conduct the act of preservation effectively at all times, regardless of the respective current mindset and the technological platform?

After these considerations, a momentous question should also be strongly posed: Why are the local research funders so quiet on this topic??

2. Closing the Generation Gap

The cultural differences between the younger generation and their parents are expressed today in such a way that we experience in the field of "new technologies in the workplace" the following remarkable fact: the parental generation, in the application of these technologies, obtains the support of their *next generation* and, consequently, often learns *from* their *next generation* while dealing with these technologies. The transfer of know-how and experience therefore occurs in persons from *younger to older* alike. At the same time, we are seeing for the first time in history, and within the last few years, how a rich, technical apparatus is created – accompanied by the respective know-how – for which the particular content literally has yet to be found. This is now being created mostly by commercial "content providers" with the benefit of hindsight. Formerly it was the reverse: first came the content (presentation of a problem), and then came the solution as a derivative of know-how. In summary: Technology today is the "form", and chronologically speaking, is first in the world, whereas only until the second instance does one wonder, "cui prodest?" Content or various sources of content then result, because they were made for this purpose. Does technology now drive methods and processes or vice-versa? Now it behaves in such a way that today's administrators of their cultural heritage have been trained at a time when there were other methods and decision-making processes in the scientific establishment, and they

[1] Italo Calvino, Lezioni americane. Sei proposte per il prossimo millennio.Garzanti, Milano, 1988

therefore often applied an apparatus that can be derived from this earlier thinking. It is not timely. These circumstances lead to a strengthening of the generation gap.

Corollary to the formation of a matrix: The methods handed down in the scientific establishment from one generation to the next as well as decision-making processes should be called into question to the extent where the following principle may come to bear: What really matters is not whether a particular system is perfect or true ("true" is an arbitrary term), but rather how well it functions for the respective user/user groups. Efficiency is the measure of this "truth".

3. Digital Archives and Cultural Heritage: Rethinking Workflows, Administration, Decision-Making Processes

The last years were marked by the so-called convergence of technologies. What was left out of the general discourse was the convergence of knowledge (or better stated: the convergence of disciplines). The focus was not, as apparently widely believed, on extreme specialization, but on the convergence of disciplines. The projects that are practiced today arise from this convergence. Digital archives can be taken as an example here. Really successful projects are solely those projects where different disciplines are integrated into one: law, computer science, linguistics, psychology, philosophy of science, communication science, economics, sociology, and so on and so forth.

Admittedly, the Internet was created in academic circles and used for the first time on a broad basis. However, there are a large number of findings, knowledge-related processes and functions that our academic communities are only just now beginning to understand on a broader basis where they in turn then very slowly process such data. Chronologically this takes place after someone else, namely the *next generation*, has already successfully used it. In this context, the ease with which one can take advantage of the terms for digital repositories offered by Calvino for the Norton Lectures, is truly amazing.

First, a brief note on the term *consistency*: here, the "shelf life of the data" is meant and the remarks on preservation can be found again throughout the entire text.

Multiplicity: This not only stands for the consistency of the data (e.g. metadata) and of work processes including the complexity of the resulting processes in the digital archive (e.g. with preservation and functions of reuse), but also the complexity of the system itself. The *next generation* has been practicing networking and the building of communities for years. The young students of today are the young scientists of tomorrow. They are the ones who cavort in the social web for years and for example, create lists of "friends" on their Facebook pages and link to multimedia content. That is exactly what they are doing, deliberately and consistently, including the citability of dataset: How many repositories would have to be made available to our traditionally run, educational scientific institutions in Europe in order to benefit – assuming academic methods – similar communities of students and their teachers involved in research?

Corollary to the formation of a matrix: It is time to create similar codified opportunities for our repositories, such as chat rooms for setting up a digital repository. Another measure would be to allow in our universities access to the functions of the systems in order to open them up also to guests (befriended scientists or partners in a project). Privileges and access policies should be the same as for the "own community".

Lightness: Lightness requires doing without the *exclusive use* of centralized logic, central systems, and "central intelligence".

In the context of digital archives, the lightness of operation and the "lightness" of the data are essential characteristics one expects from digital archives. This lightness is synonymous with

intuitive controls and should be realized by using common and generally accepted standards. A generally accepted standard does not necessarily mean "a certified standard". The "lightness of the information structured in accordance with generally accepted standards" may be the lowest common denominator of the demands of all concerned, qualified providers of data to a scientific digital archive. Defining "lowest common denominator" could result in the following task: The task of the system is the fostering of communication between those things that are different. In this case, the differences would not be blurred, but, on the contrary, strengthened by highlighting the characteristics of individual digital objects. (Regarding the digital objects: no matter what format and type, they must be provided with contextual information and equipped with technical, descriptive, and long-term digital preservation metadata at their inception).

We have seen how some platforms have prevailed worldwide. What each of these have in common is namely their lightness. I will mention four of them at random: Napster (who can still remember the peer-to-peer exchange of data and the author alert systems used then?), eBay, Amazon, and YouTube. Let's stick with YouTube for a moment and take as an example the lightness of use in uploading content. This is not just about the operation, here the focus is primarily on access to the data delivery process (*anyone* can upload content) and on access to information (*anyone* can download content). The lightness is shown in a further example: The *next generation* uses Flickr as one of the world's best online photo management and sharing applications. Why doesn't something similar happen in the scientific establishment?

Corollary to the formation of a matrix: The scientific community should have the complete opportunity to store images, pictures and powerpoint presentations in institutional repositories of their institutions quickly, easily and inexpensively (see also *multiplicity*). Here the emphasis is on the processes of targeted publishing, quoting, commenting, sharing and reuse by and with the interested public, or community.

Another corollary to the formation of a matrix: The previous corollary implies that digital content is reliably archived long-term, provided with appropriate metadata, and more easily and always searchable via a persistent signature (assignable to the digital content and/or the respective author with a digital author identificator).

Quickness: The speed of the system is determined by the normalization of data (removal of redundancy) and by dividing it into areas that are associated with a specific task. The concept of the system operator should be determined primarily by efficiency and intuitive recognition.

The efficiency of the user interface – interaction concept – should be distinguished by the fact that the information content on the screen is not too compressed. From the above platforms, I now move to eBay. The operation of the platform requires a low level of literacy, although the services offered can be very complex (e.g. the clarification of payment terms, the resolution of legal issues and questions of logistics.) The user focuses only on his projects, all collateral duties will be met by the system (e.g. allocation of tags for indexing.) The same goes for Amazon. In all of these systems used worldwide, ethnic, political, or linguistic borders are irrelevant, rather, a variant of the game of accessibility becomes effective.

Corollary to the formation of a matrix: The assignability of information should take place quickly and in one effort. The user should always know where he happens to be, what he is doing and how he can cancel a transaction. *He may never get lost in the system*. Accessibility plays an important role. Accessibility is not just a purely technical *issue* to be solved, rather, it must be a fixed part of the deliberations at the stage of conceptual planning.

Exactitude: The accuracy of the descriptive data is to be achieved through standardization and coding. The submission of data should be conducted according to a set standard, supported by an information code for the individual entries. Thereby, one or more subject catalogs should be used, which are used to classify which have equivalents in other languages and which have cross-references. The systematic accuracy should be determined by syntactic accuracy (the syntax of the user input is determined and controlled through the system in each case), and by semantic accuracy. At this point, an incidental remark made in 1984 by Calvino on "exactitude" can be quoted. For him, exactitude was three things, and for this essay, I will employ namely his second point mentioned: "The evocation of clear, distinctive and memorable visual images, in Italian we have an adjective for this that does not exist in English nor German: icastico from the Greek *eikastikós*."[2]

Here it is quite remarkable how, still in subsequent years with the *next generation*, the term "icon" could prevail in everyday language in its wide-ranging application, and now in both languages precisely this semantic aspect in English and German has become indispensable.

Corollary to the formation of a matrix: Also in this case, the effectiveness is the measure of true accuracy. In general, no "blank spaces" should exist (i.e. "null values" for information derived from queries should not be allowed.)

Visibility: In this context, visibility is considered the ease of use. The user should be able to carry out actions based on on-screen information and interact with the system, or alternatively, to retain on-screen information as a consequence of actions the user takes. Furthermore, the degree of traceability is not only significant for the objects (e.g. the origin) but also for the work processes (that is, the traceability in the search history and its results).

Corollary to the formation of a matrix: Not only the visibility of the repositories should be increased. Institutional blogs should be conducted at the interfaces to the repositories of our institutions. The digital content, the content of certified repositories, should be organized in relation to each other or linked (e.g. in order to create new collections of digital objects *and* to scientifically annotate them). Qualified content should be posted, linked and annotated. Even certainly the linking of the repository and long-preserved material should be linkable to and from platforms like, for example, Twitter. We need a cross-disciplinary approach.

4. Limits to the Digital Preservation of Cultural Heritage

What prevents us from following these corollaries? Mainly honored traditions and processes (methods), even if some or even all of the above conditions are satisfied. In addition, crucial is the lack of confidence in the know-how developed to date, as well as in the expertise of active promoters of these processes. The *next generation* (and a part of it is our young scientists) is often unfortunately not taken seriously because they are considered too young (which is also the reason why this writing does not include something about Twitter-like services nor something about the possibilities that would result from safeguarding cultural heritage for use in mobile applications. At this very instance, developed know-how in general seems methodologically too young.

Corollary to the formation of a matrix: We need to foster the building of confidence in expertise, competence and the available *e*-infrastructure. It would also be recommended to develop certification mechanisms for digital archives, so that reliability would be resultant and thus the citability of datasets would be enabled. Quality Assurance in all of its facets would then be a part of the certification mechanism.

[2] Cited from the German translation of Calvino's work, page 83, in: Italo Calvino, Sechs Vorschläge für das Nächste Jahrtausend, Harvard Vorlesungen, Carl Hanser Verlag, Munich Vienna, 1988

5. Conclusions

In safeguarding cultural heritage, we need a more sophisticated way of thinking for developing solutions. The approach to the design of systems – including systems of thought – should be crossdisciplinary. We need a crossdisciplinary approach.

The instrument itself, the digital repository, should be designed from the beginning as a multimedia marketing tool that enables information transfer and communication between users (data suppliers and consumers) in order to make content consistently available. The possibilities for distribution of information would be far more diverse and knowledge would not only be stored but its implementation would be easier. To accomplish this, a different access system would have to be designed from the beginning, including a sophisticated rights management system. Of course, the data provider should retain all sovereignty over the data. These are not empty words as the solution for these issues today is not of a technical but exclusively of a political nature. Access and restrictions (technically and legally speaking) are mostly an expression of political will. The same goes for accessibility: it is not only challenged technically, but also in all of its associated processes, in data production ranging from the delivery of data to the final data output.

We should therefore redefine the role of users. Users can, in principle, be individual users or institutions. Users can also be subdivided into groups of data providers and consumers. This necessitates a policy of open and free access not only for the consumption of published information, but also in the very *process of publishing itself*.

With regard to users, generally speaking, they should be empowered, especially that user access should be enhanced, particularly in the following two roles and processes:

1. The user as a data supplier with free access to the digital archive

2. The end-user as the beneficiary of the digital content in the digital archive.

It should be possible to guarantee this end-user free access to all published information. He should be given more rights and functionality. This requires a different approach with the wishes of the end users (focus groups) on the system. For example, for the purposes of the reuse of digital content, the end-user should be in a position to be able to implement the new knowledge gained by linking it with other content online (e.g. formation of collections of data sets inside of the repository). In addition, he could be empowered to link individual objects together so that he can therefore "form virtual dossiers" which he can them make available to other users in his community.

Interoperability with other systems and trust in online interaction should be guaranteed to the user (individual user or institution) as a data supplier. For information providers, who are not from the *next generation*, special training should take place offering these users more expertise, especially in the areas of preservation and reuse of information and techniques of self archiving.

Finally, perhaps the most important corollary on the formation of a matrix comes this time in a personal form (and please forgive the repetition):

What really matters is not whether a particular system conforms to a "true" norm ("true" is an arbitrary term), but rather how well it functions for you the user. Effectiveness is the measure of this "truth".