LONG-TERM PRESERVATION OF 3-D CULTURAL HERITAGE DATA RELATED TO ARCHITECTURAL SITES

E. Paquet [a] and H. L. Viktor[b]

[a] Visual Information Technology, National Research Council, M50 Montreal Road, Ottawa, K1A 0R6, Canada – eric.paquet@nrc-cnrc.ca.ca

[b] SITE, University of Ottawa, 800 King Edward Ave., Ottawa, K1N 6N5, Canada – hlviktor@site.uottawa.ca

Commission VI, WG V/2[a]

KEY WORDS: 3-D, Cultural Heritage, Indexing, Visualisation, and Long-term Preservation

ABSTRACT:

With the recent developments in three-dimensional (3-D) scanner technologies and photogrammetric techniques, it is now possible to acquire and create accurate models of historical and archaeological sites. In this way, unrestricted access to these sites, which is highly desirable from both a research and a cultural perspective, is provided. Through the process of virtualisation, numerous virtual collections are created. These collections must be archives, indexed and visualised over a very long period of time in order to be able to monitor and restore them as required. However, the intrinsic complexities and tremendous importance of ensuring long-term preservation and access to these collections have been widely overlooked. This neglect may lead to the creation of a so-called "Digital Rosetta Stone", where models become obsolete and the data cannot be interpreted or virtualised. This paper presents a framework for the long-term preservation of 3-D culture heritage data as well as the application thereof in monitoring, restoration and virtual access. The interplay between raw data and model is considered as well as the importance of calibration. Suitable archiving and indexing techniques are described and the issue of visualisation over a very long period of time is addressed. An approach to experimentation though detachment, migration and emulation is presented.

1. INTRODUCTION

Over the last decades, 3-D technologies have acquired a tremendous importance in cultural heritage. These technologies can provide highly valuable elements of solution to the problems of monitoring, public access and preservation. Let us review these problems briefly.

Historical and archaeological sites suffer natural degradation through time. The process of degradation can be accelerated if the site is frequently accessed, is exposed to the natural elements or if the related artefacts are repeatedly manipulated. Natural degradation is, up to a certain point, inescapable. The first action that can be taken is to isolate the site from the public. This is not, of course, in the public interest but a virtual copy of the original can replace the later and can be visualised either with the help of a virtual theatre or with a graphical workstation.

There is a clear limit to this approach. On an emotional level, most people want to see reality and not only virtuality. From a practical point of view though, the virtual copy allows, at least in principle, an unlimited number of persons to interact with the model and to manipulate the virtual artefact in many possible ways. Of course, this is not possible with the original, since it has only one instance. Also, most of the time severe restrictions are placed on the way and by whom such an artefact can be manipulated, e.g. Phoenician tablets. Even in such a case, degradation won't be prevented and some restoration, or at least consolidation, might be needed at some point in the future. The decision to proceed with a restoration is a difficult one [1]. If the decision is taken too early, the restoration might create more problems than it solves. On the other hand, if it is undertaken too late, irreversible damage to the artefact occurs. Consequently, it is important to monitor the level and the rate of deterioration to which the artefact is submitted.

Malicious acts can also provoke the degradation, if not the complete destruction, of a site. One can think of an isolated act like the attack on Michelangelo's Pieta in the Vatican Basilica, or a government decision, such as the destruction of the Buddha of Bamiyan in Afghanistan.

Figure 1. The Mostar Bridge in Bosnia and Herzegovina before the Bosnian War. The bridge was an icon of the friendship between the Muslim and Serbian communities

Cultural site are often destroyed in order to wipe out their symbolic value. For example, consider the case of the Mostar Bridge in Bosnia and Herzegovina, which was considered a symbol of friendship between the Serbian and the Bosnian community. Sites can also be destroyed by natural phenomenon such as the frescoes painted by Giotto in St-Francis Basilica in Assisi, which, were partially destroyed by an earthquake.

Figure 2. The figure shows the Mostar Bridge some time after the Bosnian War. A temporary bridge was put in place. The bridge has recently been restored and reopened.

At this point, one can feel the importance of having a precise, accurate and faithful three-dimensional digital copy of the original site. The virtual model can be used as a reference model in order to monitor the degradation of a site. The reference model encapsulates the information about the three-dimensional shape, and, in some cases, the colour distribution or textures, of the site at a certain time. By acquiring a new digital model of the site at a later stage, it is possible to compare the two virtual models in order to estimate the level of degradation. In addition, this has to be supplement with spectral and chemical analysis, just to mention a few. The sine qua non condition here is that the virtual copy must be faithful to the original, up to the required level of resolution and accuracy. This level is determined by the nature and the historical importance of the site. If the model is not faithful, the comparison of the two models does not show real differences, but rather acquisition artefacts. Consequently the models have very little use for monitoring purposes. The comparison is usually carried out by registering the two models and by comparing them according to a certain metric, such as the quadratic form.

The virtual copy may also be utilised for displaying the artefact or site of interest to the grand public. For instance, one might display the interior of an Egyptian tomb in Iceland, display a site that has been sealed, like the Grotto of Lascaux (in 1963), or to prepare the visitor for the visit of a real site for which the access is restricted. A good preparation has tremendous importance if the visit has to be done under stringent condition: for instance, if the visit is limited to a few minutes. This example shows that preservation and public access are not necessarily antagonist; at least as far as the virtual model is concerned!

Most of the above-mentioned problems as well as the related issues such as data acquisition, calibration, sensors, data registration, models creation, colours acquisition either on a per vertex basis or as a set of texture maps, colour calibration, visualisation, description and content-based indexing of the models, just to mention a few, have been extensively studied in the literature [2].

Nevertheless, one problem has received very little attention so far: the long-term preservation of three-dimensional data. This problem is of tremendous importance for many reasons. Let us review some of the issues we brought earlier. Let us suppose that a heritage site has been virtualised and that it is destroyed or badly damaged by an earthquake some hundred years later. Technically, it is possible to base the restoration on the virtual model. This is feasible if the earthquake happened a few years after the sites was virtualised, but it might be highly problematic if it happened a century later. For instance, the file format might not be in use anymore, the graphic primitives might not be supported by current visualisation software, the operating system and the format of the data might not exist anymore, to name but a few [3, 4].

The problem is that the lifetime of digital data is very short. A papyrus can easily last thousand of years if it is properly stored in a dry location, but the situation is completely different in the case of digital media. The media on which the data are stored usually has a very short lifetime and the very nature of the data tends to change or evolve dramatically over very short periods of time [5]. Some might say that this is a conceptual problem; one just has to virtualise the site or the artefact on a regular basis. As it is well known, this solution is not feasible both for practical and economical reasons.

Famous sites might be virtualised frequently but this is very unlikely for minor or less well-known sites. The case of the opera la Fenice in Venice is quite instructive from this aspect. After the opera burned in 1996, the team in charge of the restoration rapidly realised that there was not a comprehensive survey of the opera building. Unsurprisingly, neither a photogrammetric survey nor a virtualisation was ever performed. As a direct consequence, the architect in charge of the reconstitution had to scrutinise, under a microscope, historical pictures of la Fenice in order to be able to redesign many fine features of the theatre. The lesson was that not enough information or documentation was archived and most of the information that, at one point, may have been available was lost.

This paper addresses the problem of long-term preservation of three-dimensional data in the framework of cultural heritage. In particular we want to address less evident questions including the long-term relation between the raw data, the calibration and the final virtual model. The paper is organised as follows. Firstly we investigate some aspects of the acquisition that might have an effect on the long-term preservation of data. Then, we address the archival aspect of the problem. We focus on issues like data format, standard and content-based description, retrieval and long-term dynamical storage. Next, the visualisation aspect is addressed. In particular we focus our attention on colour calibration and long-term dynamical visualisation. We also describe the design of an experimental approach which will be used to further study long-term preservation issues. A conclusion follows.

2. LONG-TERM ASPECTS OF THREE-DIMENSIONAL DATA ACQUISITION

This section studies the applications and the implications of long-term preservation of three-dimensional data in the framework of cultural heritage. In order to facilitate the comprehension, the argumentation of this section shall be developed around a specific example. Nevertheless, there is no need to say, that the realm of the procedure remains entirely general.

Let us consider once more the Mostar Bridge. The name of the bridge means “old bridge”. It was finished in 1566 after nine years of construction for the Ottoman Sultan Suliman the Magnificent. The master builder was a student of the great architect Sinan.

The bridge existed for more than four centuries but was destroyed during the Bosnian Civil War, on 9 November 1993. No model of the bridge was ever created and the architects in charge of the restoration had to base their work on historical pictures and on the testimonies of the local population. Now, if a model of the bridge would have been created, the restoration process would have been greatly facilitated. However, a long delay might separate the acquisition of the model from the moment it is brought into play. Consequently, the long-term preservation of the model becomes a major issue.

Figure 3. The figure shows one of the frescoes of the Assisi Basilica before its partial destruction by an earthquake. The fresco is title “Homage by a Simple Man” and was painted by Giotto circa 1300.

In the standard approach, the reference model is archived for future use. This procedure is unsatisfactory, as explained next. A virtual model embodies and integrates the outcome of an acquisition and of a modelling process. This process contains much more information than what is reflected in the final model. The acquisition pipeline involves the acquisition of numerous geometrical views with one or different type of acquisition systems, the calibration procedure and data, the registration of the acquired views, the acquisition of colour and texture and the corresponding calibration. When the model is created, the data are exploited in a certain way, which depends on the tools and on the approach chosen. A different tool or approach might lead to a slightly different model. In addition, with the evolution of the techniques, there might exist a better way to make use of the acquired data and the calibration.

For the above-mentioned reasons, in addition to the final virtual model, a large number of other information should be archived. These include the acquired raw data, a technical description of the acquisition system and of the raw data file format, the calibration data, the calibration procedure, the equipment and the software utilised in the creation of the model, the operation system and the type of computer, amongst others. It is then possible to audit the data, assert the reference model and even create a new model if one believes that the original raw data were misused.

Another important challenge of long-term preservation concerns the interpretability of the programs created long ago. That is, it should be ensured that the virtual model can be recreated exactly as when it was archived. This can be achieved through the use of a full emulator, where the executable format of the old programs, as contained in the archive, is run on new machines [6-8].

3. ARCHIVAL, INDEXING AND RETRIEVAL: A LONG-TERM PERSPECTIVE

This section presents various issues and solutions for the long-term archival, indexing and retrieval of three-dimensional data.

It is relatively simple to preserve data over a period of a few years, but it is quite challenging to preserve them over a period of a few decades. Consequently, a large number of digital data is highly endangered due to the so-called silent obsolescence of data formats, software and hardware [9, 10]. This serious problem is, of course, relevant not only for cultural heritage data, but also for electronic publications and other digital documents [3, 4].

There are many reasons for this difficulty. One of the first reasons is the quasi absence of standards. For instance, each brand of scanner corresponds to a different raw data format. The same happens to the final model, which, can be in various formats. There are very few international standards; as a matter of fact, most formats are proprietary. This is especially true for raw data formats, which are very volatile. The lifetime of a format is determined by the lifetime of the underlying corporation, which, in many cases, spans only a few years. The same problem applies to the related software. According to [5], the technologies for, for example, data presentation changes every four year. As a result, one should not only archive the raw data, but also the software and a complete documentation of the procedure that was used in the creation of the virtual model.

As we said earlier, one of the main interests of having the raw data is to be able to process it with more recent algorithms and techniques. In this way, the faithfulness of the model may be increased as technology advances. This can be very difficult if the formats in which the raw data are archived are obsolete. One approach is to convert the format to a new format but that can be difficult if the format is proprietary and has been inactive for more than a decade. In this case, the gathering and analysis of documentation and configurations of outdated material are very time consuming and require extensive resources [3].

An alternative is to propagate the data over time, through detachment and migration [3, 5]. This means that, on a regular basis, the data is detached from their original media. Subsequently, the original data format is reconverted into a more recent format and stored using new software and/or hardware, thus utilizing more recent technologies. This procedure is carried out at regular intervals, to ensure that the most recent technologies are being utilized.

In essence, the process of building a persistent store for the long-term preservation of data consists of two main steps, namely firstly, archiving and indexing, and secondly, retrieval. These steps follow a cycle, to be used when migrating data over time as technology changes [5]. These changes can occur on a system-level (e.g. software or scanner technology changes) or on an information model (e.g. data format or collection practices changes) level.

The benefit of such an approach is that conversion tools are available when a format is alive while they might be very difficult to find otherwise. Nevertheless, it is important to assert that the reformatting of the data does not affect their intrinsic values. When migrating the data forward, it is therefore important to ensure that the data formats and programs are also brought forward. That is, it has to be ensured that original the data collection can be recreated as and when required [9, 10]. This can be achieved through the use of an emulator, which can be used to run the old programs and to present the virtual models as seen in the past [6, 7].

From a cultural heritage point of view, it is desirable to archive the models and the raw data in a central repository. This central repository should be constructed in order to provide a high level of protection against ambient conditions, robbery and hazards like fires, earthquakes, etc. That means the repository might eventually host a substantial number of sites and artefacts. In particular, the use of a data warehouse which is based on a dimensional model is proposed here. A data warehouse stores historic data that has been collected over time and are thus ideal to store the archival data [11]. Section 5 will introduce a dimensional model for detaching, migrating and emulating the long-term preservation of 3-D data.

When employing a data warehouse, it becomes important to index the sites and artefacts in order to be able to retrieve not only the site or the artefact of interest but also the sites that might be related to a given site. The index or description can be textual, numerical or content-based. A textual description might be a text with archaeological and historical data about the artefact, a numerical description might be a spectral or a chemical analysis, while a content-based description might be a three-dimensional shape descriptor or a texture description.

Alphanumerical descriptions are relatively well known and we refer the reader to the literature. We describe two content-based descriptors: one for the texture and the other one for the shape [12].

We now depict our algorithm for texture description. The colour distribution of each image is describes in terms of hue and saturation. This colour space imitates many characteristics of the human visual system. The hue corresponds to our intuition of colour e.g. red, green or blue while saturation corresponds to the colour strength e.g. light red or deep red.

Next, a set of points is sampled from the image. A quasi-random sequence generates the points. In the present implementation, the Sobol sequence has been selected. Each point of this sequence becomes the centre of a small rectangular window on the texture. For each centre position, the pixels inside the corresponding window are extracted and the associated hue and saturation maps are calculated.

Figure 4. Texture retrieval from the palace of Knossos in Crete. The two textures present a similarity of palette and style.

The statistical distribution of the colours within the window is characterized by a bidimensional histogram. The first dimension of this histogram corresponds to the hue or the saturation quantified on a discrete and finite number of channels. The second dimension corresponds to the relative proportion of each channel within the window. This bidimensional histogram is computed and accumulated for each point of the sequence, i.e. the current histogram is the sum of the histograms at the current and at the previous position. From this process, a compact descriptor or index is obtained.

This index provides an abstract description of the composition of the texture, i.e. of the local distribution of colours throughout the image. This is very important. This index does not represent a global description of the texture nor is it based on a particular segmentation scheme. Instead, it characterized the statistics of colour distribution within a small region that is moved randomly over the image. Consequently, there are no formal relations between the different regions. This implies that the different components of a texture can be combined in various ways, but still be identified as the same texture. For instance, consider a floral pattern appears that at a different position. For this reason, the algorithm is robust against occlusion, composition, partial view and viewpoint. Nevertheless, this approach provides a good level of discrimination. Content-based indexing of images is important to ensure the long-term preservation of data, since it provides a way by which unorganised data can be indexed and searched automatically. Such data are very common in practical applications, e.g. consider a collection of photographs. It follows that, if a specific image cannot be found because of obsolete, outdated indexing, future access to it may be impossible.

The algorithm for three-dimensional shape description can be outlined as follows [12]. The centre of mass of the artefact is calculated and the coordinates of its vertices are normalised relatively to the position of its centre of mass. Then, the tensor of inertia of the artefact is calculated. This tensor is a 3 x 3 matrix. In order to take into account the tessellation in the computation of these quantities, we do not utilise the vertices per se but the centres of mass of the corresponding triangles; the so-called tri-centres. In all subsequent calculations, the coordinates of each tri-centre are weighted with the area of their corresponding triangle. The later is being normalised by the total area of the artefact, i.e. with the sum of the area of all triangles. In this way, the calculation can be made robust against tessellation, which means that the index is not dependent on the method by which the artefact was virtualised: a sine qua non condition for real world applications.

Figure 5. Shape retrieval is suitable in the case there is a homomorphism in between the artefacts. From one reference artefact or archetype it is possible to retrieve related and similar items base on three-dimensional shape, appearance and style; for instance Corinthian columns. The two vases illustrated are made of stone and are from the so-called Old Egyptian Kingdom circa 2700-2200 BC.

In order to achieve rotation invariance, the Eigen vectors of the tensor of inertia are calculated. Once normalised, the unit vectors define a unique reference frame, which is independent on the pose and the scale of the corresponding artefact: the so-called Eigen frame. The unit vectors are identified by their corresponding Eigen values.

The descriptor is based on the concept of a cord. A cord is a vector that originates from the centre of mass of the artefact and that terminates on a given tri-centre. The coordinates of the cords are calculated in the Eigen reference frame in cosine coordinates. The cosine coordinates consist of two cosine directions and a spherical radius. The cosine directions are defined in relation with the two unit vectors associated with the smallest Eigen values i.e. the direction along witch the artefact presents the maximum spatial extension. In other words, the cosine directions are the angles between the cords and the unit vectors. The radius of the cords are normalised relatively to the median distance in between the tri-centres and the centre of mass in order to be scale invariant. It should be noticed that the normalisation is not performed relatively to the maximum distance in between the tri-centres and the centre of mass in order to achieve robustness against outliers or extraordinary tri-centres. From that point of view, the median is more efficient that the average. The cords are also weighted in terms of the area of the corresponding triangles; the later being normalised in terms of the total area of the artefact.

The statistical distribution of the cords is described in terms of three histograms. The first histogram described the distribution of the cosine directions associated to the unit vector associated with the smallest Eigen value. The second histogram described the distribution of the cosine directions associated with the unit vector associated with the second smallest Eigen value. The third histogram described the distribution of the normalised spherical radius as defined in the previous paragraph. The ensemble of the three histograms constitutes the shape index of the corresponding artefact. From a long-term preservation perspective, the content-based indexing of 3-D shapes provides a way by which unorganised data can be indexed and searched automatically. For example, consider a collection of 3-D models which need to be searched to retrieve a specific model. A specific model that cannot be found, due to poor indexing, may be lost for future use.

The above-mentioned descriptors allow an automatic description of the appearance and the shape of an artefact. They can be use to search directly by shape and appearance, or in order to compensate for the absence of alphanumerical description.

4. VISUALISATION OF THREE-DIMENSIONAL MODEL OVER A VERY LONG PERIOD OF TIME

In this section, issues related to visualisation of three-dimensional model over a very long period of time are reviewed and addressed. Four particular aspects retain our attention: the display of data, the visual colours calibration, virtual restoration and virtual tour.

As we saw, digital file formats usually have a very short lifetime. In order to ensure that the data are available in the distant future, they must be migrated or propagated in time. That means that the files have to be transformed into a new format, in other to ensure that they are compatible with the current devices and software. As mentioned earlier, this process can be potentially harmful for the data. That is, some data or some precision might be lost in the process. Consequently, if it is impossible to find a format that can encapsulate all the information present in the previous one, it is therefore better to preserve the original data in the prospect that a better match could be found in the future.

The conversion of three-dimensional information may be relatively easy if we assume that the model is saved as a mesh or as a set of points, each point represents a measurement in three-dimension. This constitutes an unambiguous body of information in the sense that X mm today shall be X mm Y years from now.

The textures present a particular problem in terms of calibration. Most popular texture formats such as JPEG do not encapsulate any calibration information [8]. Despite the fact that the colours might be acquired with a calibrated apparatus, for instance with a white beam laser scanner for which a correction in terms of the angles of incidence and reflection is applied, the texture format still just provide a set of numbers. Such a set of numbers has no real optical meaning without the calibration. The situation becomes more acute if one wants to display the data. This is due to the fact that the calibration, if any, of the visualisation system might not correspond to the calibration of the acquired data. Most of the time, these variations are relatively small when employing "state of the art" monitors. However, these small calibrations can be very significant. The point is that even small variations are important in cultural heritage applications. In particular, such small variations may be utilised to monitor the deterioration of a site over a certain period of time. That means that one has to compare the texture maps: such a procedure is meaningful if and only if the two maps present the same calibration.

In order to achieve this, the calibration procedure and the physical calibration device must be archived with the original files. Such approach has been applied with success. For instance, when the Eiffel Tower was repainted, the French Government decided to use the original colour. Neither the company nor the paint existed anymore, but a calibration device called a "nuancier" was kept. From this device, it was relatively easy to recover the original colour.

Figure 6. A 19th century painting by Panini showing the interior of the Pantheon in Rome. The scene was painted according the laws of perspective.

Even if the colours are acquired with the same calibration, the visualisation problem is not entirely solved. The display must be calibrated in order to render the colours with the same calibration that was employed during their acquisition. Again, it is important to have the original calibration device in order to calibrate the display, i.e. to ensure the colours are reproduced as faithfully as possible.

The three-dimensional aspect of the data is not as problematic in the sense that a three-dimensional point is an absolute measurement. Nevertheless, some caution must be exerted when displaying a three-dimensional site in stereo. The display must be calibrated in order to reproduce the physical condition of visualisation. If this is not the case, the model might be perceived as slightly distorted. Subsequently, false conclusions might be drawn from a visual inspection of the virtual model.

With a long-term perspective, displays have two very important domains of application: virtual restoration and virtual tour. When a site is damaged or destroyed, it is important to study the various scenarios of restoration before proceeding to the restoration per se. Various actions can be performed on the virtual model in a proper virtual environment. For instance, one can virtually remove a part that was added at a later time, clean the paint, restore some frescoes and evaluate the impact of a specific action on a site.

The virtual visit is also an important aspect of the long-term preservation of data. We have already mentioned, in the introduction, that a virtual visit can replace the effective visit of the site if the later is closed or destroyed. It can also be used as a preparation for a visit that, for conservation reasons, is required to be short. But there is another aspect, which in our opinion has been neglected: the evolution of the site.

Let us consider a site that has been monitored on a regular basis and for which a virtual model is created after each acquisition. If the monitoring is performed for a sufficiently long period of time, the site might go through various phases of alteration due to deterioration, restoration, modifications and so on. An instructive example is provided by the Pantheon in Rome, which is well documented both in terms of pictures and paintings. Many archived photographs and paintings are available, from which is would be possible to create a virtual model of the Pantheon at different moments in its history.

The case of paintings is of particular interest. Many of the works of art were created in strict accordance with the laws of perspective. That means that it is possible to recreate a virtual model from a single painting by enforcing the geometrical constraints present in the scene with a bundle adjustment technique. These constraints can be, for instance, that some columns have the same heights, or that some lines that appear in the scene are either parallel or perpendicular and so on.

Virtual displays provide us with the means to visualise the evolution of the site through time and to transform what is, by definition, a static model into a dynamic and lively historical evolution.

5. EXPERIMENTAL DESIGN

The development of methods for the archiving, indexing, retrieval and visualisation over a very long period of time poses an important research question. This section discusses the design and implementation of a system to address the issues discussed in earlier sections, focussing on the detachment and migration of data residing in a persistent data warehouse, combined with the emulation of the virtual models.

To this end, we present an experimental approach for three-dimensional data preservation of cultural heritage sites. During this experimentation, issues such as scalability of information and infrastructure, managing heterogonous data sources, ensuring high quality calibration information storage and handling updating of hardware and software, emulation of virtual models, amongst others, will be further investigated.

In this research, a number of virtual models will be archived using an IBM DB2 data warehouse and ultra-fast fibre-optic mass storage, hosted at the Intelligent Data Lab of the University of Ottawa. The retrieval and visualisation components will be based on indexing and visualisation technologies developed at the National Research Council of Canada [12-13].

Figure 7 shows the proposed life cycle architecture for the long-term preservation of 3-D data. The figure shows that first step, after the acquisition of the data, is indexing, as discussed in Section 3. Here, “data” refers to the data describing the body of knowledge regarding the site, including the raw data, the documentation, the virtual model, the calibration information, the programs used, etc.

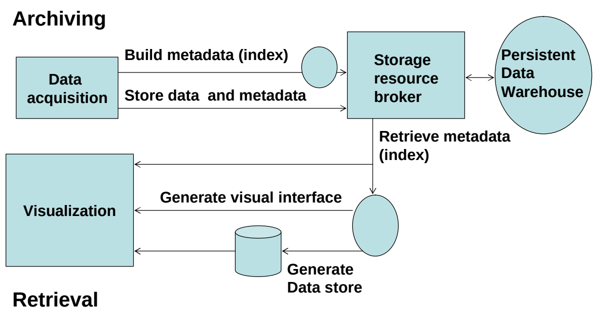

Figure 7: Lifecycle architecture for the preservation of 3-D data (Adapted from [5])

Firstly, the data are detached from the original media, in order to ensure independence from proprietary sources. The resultant data is then migrated into a persistent data store or warehouse. As we noted above, care has to be take to ensure that migration will not degrade the original information. This process will be verified by performing cyclic multiple migrations, i.e. to verify that the resultant is the same as the original [8]. A storage resource broker is used to ensure that the migration process is transparent.

In order to compare previous virtual models of a site, we propose an approach to retain the previous virtualizations through emulation [6]. Here, the executable programs running on yesterday’s machine is stored, together with an emulator of the original machine. Of course, this emulator needs to be written while the machine is still available [7, 8]. In this way, the old software can be executed and the virtual model is thus seen as in the past. This process will be verified by comparing the visualisation on the old machine with that of the emulator. Through this process, the user is therefore able to use the old visualisation and to compare it with the results obtained through newer technologies.

A similar observation can be made about the modelling process. The emulator permits to audit the original raw data with the same software that was utilised for the creation of the original model and to reconstruct the model from the original raw data in order to insure the quality of the original modelling. It is also possible to obtain a more accurate version of the model by reprocessing the raw data with more recent and powerful modelling techniques.

When retrieving data from the archive, the system provides three options. In the first option, the data is directly accessed and visualised using the current technologies as retrieved from the data store. Secondly, a new visual interface is created and the data is subsequently visualised. This option is useful when the visualisation module has been updated. The third option involves the generation of a new data mart or warehouse, to facilitate database management system software changes. It follows that these changes should be propagated to the persistent data warehouse, or on the very long term, be used to replace the original data warehouse. This process is cyclic, and it follows that new technologies will be archived continuous.

The persistent data store or warehouse will be based on the dimensional model. Dimensional modelling provides an intuitive structure to present data that has been collected over time [10]. In this approach, the time and date recorded are explicitly included as part of the data. The dimensional model for the data warehouse is shown in figure 8. The data itself is stored in different dimensions (or tables), which are linked via a central *long-term preservation fact* table. This model allows for fast, intuitive and easy querying to, for example, compare different models of a site as acquired over time.

Note that, for each of the dimensions, numerous attributes are collected. For example, the calibration dimension will include the original calibration, the calibration procedure, the calibration of the colours and so on. The virtual model dimension records the geometrical model, a description of the format, the colour vertices and the texture maps, etc. The scanner information will be stored in the hardware dimension and will include details such as a technical description of the scanner, its precision and the accuracy, the resolution, amongst others. The date dimension details the day, month, year, season, historic context, etc., in order to provide an information rich repository. The data warehouse will also include documentation regarding the software used, and in particular the procedures followed, together with the source code, as needed.

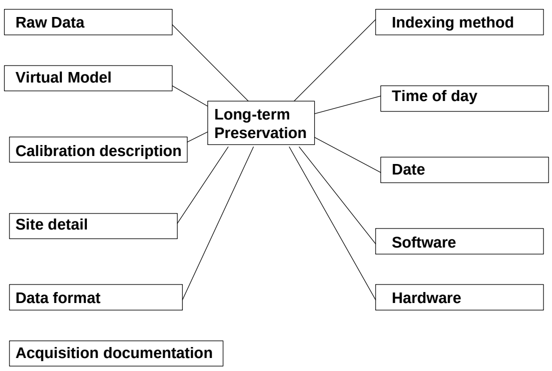

Figure 8: Dimensional model of 3-D long-term preservation data warehouse

For example, consider the Mostar Bridge introduced earlier. We assume that there have been two virtual models build of this bridge, one in 1980 and one in 1993. For each one of these models, the raw data together with details such as the scanner technologies, indexing approach, virtualisation method, calibration method and equipment used, software used, etc. were archived, using the detachment and migration approach as discussed above. This data can now be retrieved and used to visualise the two models. Let us assume that, for these two virtual models, the visualisation software used in 1980 and 1993, respectively, have been stored. Emulators are subsequently used to visualise these models, as it would have run against the original old machines. In addition, the raw data, calibration information and site detail are also used to visualise the model using the most up to date technology. In this way, the virtual models are shown not only using current technologies, but also as viewed via the original visualisation.

A number of research issues will be further investigated. This includes, but is not limited to, the following: Storing adequate information regarding calibration, managing heterogonous data formats, diverse indexing methods and virtual models, handling new software and scanner technologies, emulation of old technologies, to name but a few. Another important issue which will be addresses is the scalability of the information and the storage infrastructure, as the data warehouse expands over time. The management of the detachment and migration of the data as well as the emulation of the virtual models indicate to be a major task.

6. CONCLUSIONS

The long-term preservation of digital data poses new challenges to the research community. Conservation efforts regarding the preservation of electronic digital documents involve organizations such as UNESCO, government agencies and research groups, who are concerned with the threat posed by the "Digital Rosetta Stone" [3, 5, 10, 14, 15]. However, the authors are unaware of any project which addresses the long-term preservation of the 3-D data obtained from historic and archaeological sites in a central persistent data repository.

A framework for the long-term preservation of 3-D cultural heritage data as well as their applications in monitoring, restoration and virtual access has been presented. We stressed the importance of preserving not only the final model but the raw data, the calibration, the methodology, the software and the hardware.

The paper also presented an approach to experimentation through detachment, migration and emulation. We presented a life-cycle architecture for the long-term preservation and the dimensional model of a persistent data warehouse to be used to archive, index and retrieve the three-dimensional data.

7. REFERENCES

1. Kuman, S. et al., 2003. Digital Preservation of Ancient Cuneiform Tablets Using 3D Scanning. In: *Fourth International Conference on 3-D Digital Imaging and Modeling*, Ottawa, Canada, pp. 326-333.

2. Beraldin, J.-A. et al., 2004. Virtual Heritage: The Cases of the Byzantine Crypt of Santa Cristina and Temple C of Selinunte. In: *10th International Conference of Virtual Systems and Multimedia – VSMM 2004*, Ogaki City, Japan, pp. 32-39.

3. Rodig, P. et al., 2003. Preservation of Digital Publications: An OAIS Extension and Implementation, *Proceedings of the 2003 ACM Symposium on Document Engineering*, Grenoble: France, pp. 131-139.

4. Bergeron, B., 2001. Dark Ages II – When the Digital Data Die, Pearson Education.

5. Rajasekar, A. et al., 1999. Collection-based Persistent Archives, Proceedings of the 16th IEEE Symposium on Mass Storage, March 1999.

6. Rotherberg, J., Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, CLIR, 1999.

7. Lorie, R., A Methodology and System to Preserve Digital Data, Joint Conference on Digital Libraries, ACM/IEEE, Portland, 2002.

8. van Wijngaarden, H. and Oltmans, E., Digital Preservation and Permanent Access: the UVC for Images, IS&T Archiving Conference, San Antonio: USA, April 2004.

9. Ronchi, A.M., 2004. From Hammurabi Codex to Rosetta Stone: Long term Preservation of Digital Archives, Medici Framework Panel, 2004 World Wide Web Conference, New York, pp.1-15.

10. Heminger, A.R. and Robertson, S.B., 2000. The Digital Rosetta Stone: A Model for Maintaining Long Term Access to Static Digital Documents, Communications of the Association for Information Systems, 3 (2).

11. Kimball, R. and Ross, M., The Data Warehouse Toolkit: Dimensional Modeling, 2nd Edition, John Wiley and Sons, Canada, 2002.

12. Paquet, E. et al., 2003. Virtualization, Virtual Environments and Content-based Retrieval of Three-Dimensional Information for Cultural Applications. In: *IS&T/SPIE's 13th Annual Symposium on Electronic Imaging: Science and Technology* – Videometric, Santa Clara, USA, pp. 137-147.

13. Paquet, E. and Viktor, H.L., 2004. Visualisation, Exploration and Characterization of Virtual Collections. In: *TS W/G V/6 Visualization and Animations, XXth Congress – International Society for Photogrammetry and Remote Sensing – ISPRS 2004*, Istanbul, Turkey, pp. 597-602.

14. UNESCO. Preliminary Draft Charter on the Preservation of the Digital Heritage, http://www.unesco.org/webworld/ica_sio/docs/28session/annex5.rtf

15. CSDS. Reference Model for an Open Archival Information System (OAIS). Technical report, Consultation Committee for Space Data Systems (CCSDS), 2001.