A DELPHI ASSESSMENT OF THE
DIGITAL ROSETTA STONE MODEL

Alan R. Heminger, Ph.D.  Associate Professor
Air Force Institute of Technology
alan.heminger@afit.edu

Don M. Kelley, Captain, USAF

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

Abstract

Information that is stored digitally can only be used if it can be retrieved and interpreted. If the methods to retrieve the bitstream are lost, it may be difficult, if not impossible to re-create them. The knowledge to interpret the bitstream is also at risk. The Digital Rosetta Stone (DRS) Model was developed as a framework for capturing and maintaining the methods necessary to retrieve and display digital information stored on obsolete media or using obsolete software. However, this conceptual model had not yet been assessed by the community of professionals for its practical efficacy. This study used the Delphi Method to explore the DRS with those responsible for maintaining access to digital data. Overall, the Delphi group expressed concerns about the practicality of developing the DRS, but agreed that it is an important concept that should be explored further. If found to be technologically feasible and economically desirable, the DRS could well lead to a long-term solution for recovering information that would otherwise be impossible to recover.

1. Background

This study addresses accessing information in computing devices that is many generations behind the current technology. People have been storing information since the dawn of human history. However, only relatively lately have people begun to store information in digital format. What makes this important is that for the first time we are beginning to store much of our historically important information in a way that cannot be read without specific, often esoteric, technologies that we may well lose.

Today, more information is being stored digitally than was thought possible even a few years ago. According to Lyman and Varian, world production of unique information has been estimated to be between one to two exabytes ($2^{60}$) per year. One terabyte ($2^{40}$) is the equivalent of printing about 50,000 trees worth of paper. One petabyte ($2^{50}$), equals the storage of about half of all United States Academic Research Libraries.

NARA was created as a repository for government documents and other historically significant materials. In accordance with 44 USC § 3102, the head of each Federal agency is charged with cooperating with NARA in the “selection and utilization of equipment and supplies associated with records.” The Archivist of the United States, in turn, is required to accept “sufficiently historical or otherwise valuable records” (44 USC § 2107). However, “NARA faces increasingly enormous quantities of records” [2].

As if sheer volume is not enough of a problem, NARA is also receiving “an increasingly diverse load of [digital] information” created using a wide variety of software and stored in a “bewildering variety of media” [21]. This predisposes information to the threat of being permanently lost, even if it is under NARA’s watchful eye. The Archivist of the United States put it eloquently when he said, “It will be worse than sad if the marvelous technologies that are giving us a new information age outrun our ability to keep a record of it” [2].

The increasing pace of technological development also creates accelerating technological obsolescence. It is this accelerating obsolescence which is threatening our knowledge of methods used to retrieve and properly display our stored digital history [21] [4] [10] [3] [2]. Accelerating obsolescence is the quickly shortening lifespan of widely used storage and computing devices (including both hardware and software). The lifespan shortening occurs simply because the time between technological generations is becoming shorter. The impetus for this race occurs when companies (both producing and consuming), get caught up in a paradigm of needing to push the technological envelope in order to stay competitive.
Many [7] [13] [11] see an accelerated rate of increase in this race. Hedstrom [6] suggests that accelerating technological obsolescence poses the greatest danger to our digital history. Even if a decision is made to move all stored information to a new storage technological generation, we simply cannot move all of our data to a current form in a current storage environment before the next generation of technology makes the current one obsolete. As one systems librarian puts it, “All those state-of-the-art machines, software packages, and compression techniques seem old before the boxes and shrink-wrap even hit the landfill” [16]. Regardless of the chosen solution, the problem is compounded by an exponentially growing data set.

Limited resources preclude efforts to preserve all available information objects. Document retention schedules set up by the U.S. Government recognize the value of maintaining legal documents for long periods of time. However, documents deemed less significant are usually destroyed or kept for only a short while. The criteria for determining document value can change depending on many and varied factors such as world events or high-profile lawsuits. This can lead to information being lost that is later deemed important but unrecoverable. In trying to upgrade some of MIT’s archival data, Zuzga [24] realized that there were some tapes that, although not currently valuable, may be of extraordinary value in the future.

The sad reality is that the overlooked consequences are already cropping up—we have irretrievably lost critical data on more than a few occasions [19]. Due to negligence, mishandling, and technological obsolescence, most of Canada’s early recordings of feature films, radio broadcasts, and video are forever lost [14].

2. Searching for solutions

A number of people have proposed strategies to address the threat to our stored digital information [23] [20] [12]. NARA’s and other’s efforts to date have been labor intensive and expensive. However, there is also a lack of agreement in the community as to which is the best way to proceed [8].

Several methods—migration, refreshing, technology museums, etc—have been proposed by different researchers [23] [20] [12]. Unfortunately, there is no agreed-upon single strategy that will satisfy all of the above criteria [8] [16]. A survey of some of the methods that have been designed for long-term access is presented here.

2.1. Media-Based Approaches

2.1.1. Printed Hard Copy On Paper. Because paper tends to be more stable than magnetic, optical, and other electronic media [10], some have proposed printing onto paper any of the information to be saved. A major flaw to printing everything out is that there is simply too much to print. “Printed documents of all kinds comprise only .003% of the total…” of information that is produced every year [11]. Another flaw is that printing the information explicitly puts it in a non-electronic medium that is time consuming to copy, manage, store, etc. The storage requirements for the roughly 250 megabytes worth of unique data for every man, woman, and child on the Earth would require 50 billion trees per year [22].

2.1.2. Micrographics. This solution, promoted by Willis [23] seeks to remedy the problem of requiring vast amounts of paper resources. This strategy is similar to printing information on paper except the medium is plastic and the information is miniaturized. Some of the major advantages are that it is already used as an archival medium with well-documented standards; it is easy to read the medium; and it can store a high resolution of detail [ibid.]. There are some disadvantages however. The medium has to be physically handled to access the information. It can become scratched in storage or use. The copy quality is degenerative (it loses about ten percent of resolution). There are also some problems with the transfer process [ibid.]. Items “born digital” are those things that are created electronically and may not be directly transferable to microfilm. These digitally born objects may include video, audio, and databases as well as many other object types.

2.1.3. Nickel Slugs. Because paper and plastic tend to deteriorate when handled and are subject to limited environmental conditions, some have suggested engraving the digital information on nickel slugs [20] [15]. This, as well as any other method to transfer the information to non-electronic media, makes it far more difficult to access the information. Since the information is stored by engraving the information, these metal slugs offer the unique attribute of lasting for thousands of years. The storage capacity for the HD-ROM, produced by Norsam Technologies, Inc., is 200 gigabytes per disc and expandable to the petabyte range. At this rate though, it would take 10,000 of these discs just to preserve one year’s worth of data. Not all strategies are medium dependent however.

2.2. Standards-Based Approaches

Several strategies regarding a standards-based solution have been posited. The major tenet of each of these strategies is that it is easier to maintain access to data if only one standard or a few standards are used.
The Universal Preservation Format (UPF) was proposed by David MacCarn at the WGBH Education Foundation in Boston. The UPF is designed to reduce the confusion caused by the “veritable explosion of formats” [12]. It also “specifies that machine-independent algorithms be encapsulated within the stored media. Two strategies, the Bento Specification and the Open Media Format “are both media technologies that approach the UPF concept” [12]. The major disadvantage of using a single format for storing all digital information is that “no computer technical standards have yet shown any likelihood of lasting forever--indeed most have become completely obsolete within a couple of software generations” [1]. The Time Capsule File System, proposed by Zuzga [24], is a similar approach to the Universal Preservation Format. It specifies a format that is “very similar to the RFC-822 format used for electronic mail” [24]. It suffers from the same drawback as the UPF in that no single standard is likely to apply to technologies developed in twenty years and beyond. Some time into the future, scientists may find ways to process information based on an octal number system instead of binary.

2.3. Other Approaches

2.3.1. Technology Museums. Another approach is to store every generation of technology, keep the machines in working order, and run them with skilled operators. This is referred to as a technology museum. This approach would benefit by extending the longevity of computer systems and their original software to keep documents readable [20]. “Because originals are so important, [NARA has] a kind of museum of equipment that will work or can be modified to work” [2]. This works on a small scale and for a short time, but for MIT which has a tremendous amount of information, the museum concept has not been as successful as MIT had hoped [24]. A disadvantage of a technology museum is that “the hardware and software for digital media change so rapidly that it would be impossible to keep an up-to-date … museum” [16]. Bearman [1] agrees with Rothenberg [20] that there are problems with technology museums. In fact, Rothenberg argues that “even if obsolete computers were stored carefully, maintained religiously, and never used, aging processes such as [metal migration and dopant diffusion] would eventually render them inoperative; using them routinely to access obsolete digital documents would undoubtedly accelerate their demise.”

2.3.2. Refreshing. This strategy is the one that is probably most often employed [16]. It “involves transferring digital materials to a new medium, for instance, changing from 5 ¼-inch floppies to CD-ROM, or from CD-ROM to DVD” [ibid]. NARA has procedures in place for refreshing. “Whenever any of the digital media in our custody shows signs of deterioration, or whenever they reach 10 years of age, we recopy the records to new media [2]. While this approach addresses the media instability problem, it does not fundamentally address formatting problems [10]. This method ensures that we will be able to access, for example, a WordStar version 1 document, but without software to interpret the document, the file will be useless.

There is a major problem with refreshing as a best strategy, however. “Disk capacity has improved 1,000 fold in the last 15 years, consistent with Moore's Law, but the transfer rate MBps has improved only 40x in the same time” [5]. This means that disk capacity has a growth rate of 25:1 when compared with the transfer rate of data. The effect of this phenomenon is that our ability to store information is far exceeding our ability to transfer it to the next technological generation.

2.3.3. Migration. Migration involves updating the format of the old digital object into what is currently used. Returning to the WordStar 1.0 document example, migration involves translating the information and storing it in Word 2000 format. This method is used frequently but Rothenberg [20] points out that that trying to undertake such a strategy results in labor and time consuming tasks that may still result in less than complete conversion. In other words, migrating all stored data to each new generation becomes increasingly infeasible, introduces the possibility for new losses, and quickly borders on the impossible.

2.3.4. The Hybrid Approach. This strategy, proposed by Don Willis among others, suggests that for information that is not “born digital”, preserving both the electronic version and a micrographic version mitigates the disadvantages of each individual method [23]. This approach is not without its own, unique drawbacks. It would, in essence, triple the amount of already exponentially-growing information—one set of information would be the original, the second set the digital copy, and the third set the micrographic copy. Using standard compression methods, there are still about 240 terabytes of printed information yearly [11]. Even though this is a tremendous amount of information, it is a tiny amount of the total information produced yearly.

2.3.5. Encapsulation. Rothenberg [20] proposes that metadata and other information be encapsulated, or stored with, the digital information object. The other information would include the original executable software and operating system along with any other pertinent data files. One factor that is both an advantage and a disadvantage is the inclusion of the software. On one hand, the encapsulated digital object
would have the appropriate software to access the data. On the other hand, because every digital object would require its own copy of the software and operating system, it would require as many instances of the software as digital objects—even if all of the digital objects were at a single repository. Current operating systems require several hundred megabytes worth of hard drive space, and typical digital object software also requires hundreds of. Storing this nearly half of a gigabyte for one file that can range in the 10’s of kilobytes seems inefficient. With the exponential data growth, maintaining individual copies of massive software sets seems infeasible.

2.3.5. Emulation. This strategy uses hardware and software emulators to access information stored on obsolete media and in obsolete formats. Rothenberg [20] suggests that there are several benefits to the approach and argues that this is the best, if not the only, approach. The first benefit is that the emulators only need to be developed one time. However, this brings up the question, “What happens when the system the emulator runs on becomes obsolete?” A second benefit is that using the emulation of original software and hardware is the only way to accurately recreate the original digital environment. This will give the digital information object the same “look and feel” as it would have appeared using the obsolete technology. The negative aspect is that each emulation would also have to be maintained. Zuzga [24] argues that there would be a “serious continuing cost” if emulation was used. One or more of these or some other strategies may well be put in place to help retain our digital heritage. However, even with this, there will be a need to recover stranded digital information.

2.3.6. Capturing metadata. Another strategy to knowledge preservation is to capture the information about storage devices and software algorithms when it is readily available, and use that information later to recover and display digital documents. Generally, this means to collect the meta-information about the hardware and software used to create a document while the storage device and software are still in general use. If one waits until the technology is no longer current, it may be too late. For example, when trying to recover the information stored on an 8-Track Punched Paper Tape, Robertson [19] was not able to map out the entire storage technique, even though the technology was relatively recent and the company that designed the technique was still a current industry leader. If the relatively simple task of being able to read paper tape is difficult, then the task of reading magnetic and other high technology media without the proper know-how is truly daunting. Now imagine the problem if it occurs 100 years later.

3. The Digital Rosetta Stone

Robertson [19] explored the long-term access problem and suggested an approach to retrieving and interpreting data stored on obsolete media, based on capturing the metadata. He called this model the Digital Rosetta Stone (DRS). Robertson's model was conceptual in nature and did not include details of how best to implement it. This study will start with Robertson's model, and using the Delphi Technique to gather information from experts in the field, will explore the feasibility of this model, and add detail to its conceptual framework.

3.1. Focus of the DRS

Recognizing that many strategies for systemic knowledge preservation are likely to be employed by many groups, the DRS is designed for those instances where those other strategies fail. The DRS does not attempt to address recovery of information from media that has degraded beyond the point being readable. The DRS was designed to be a last-ditch effort to recover stranded information, on storage devices where the bitstream is present, but the knowledge of either how to recover the bitstream or interpret the bitstream has been lost. It is therefore to be used as a digital archaeology tool—recovering information that, until now, has been beyond reach.

3.2. DRS Components

The DRS is composed of three major components that are necessary to access digital information stored on obsolete storage devices or in obsolete formats—knowledge preservation, data recovery, and document reconstruction [19]. Developing each of these processes accurately is critical to the success of the DRS. The first major process, knowledge preservation, is addressed by the Metaknowledge Archive.

3.2.1. Metaknowledge Archive. Robertson [19] proposed developing a repository of information necessary to both recover the data and reconstruct the document, which he calls a Metaknowledge Archive (MKA). This archive would be created through the act of knowledge preservation and would form the foundation for the other processes of the DRS Model. In fact, without this MKA, a file stored on an obsolete medium and/or in an obsolete format would be completely useless, even if the bits were preserved [24] [21]. Lyman and Besser [10] point out that we know the most about a digital object at the time we create it, and we can’t know for sure how we will want/need to access it in the future. Therefore, we should capture as much information about the digital object as we can at the time of creation.
3.2.2. Recovering the Bitstream. Armed with the knowledge of storage techniques, recovery technicians can begin. Data recovery is the process of retrieving the bitstream from the outdated and obsolete medium and moving it to a current storage device. If necessary, the information in the MKA could be used to create a new medium access device. The access method may be altogether different than the original device used. For instance, instead of building a CD-ROM drive to recover a bitstream, the DRSO workers might use a high-resolution scan of the CD and software to interpret the image [19]. This may help if the media is fragile and may not survive traditional data access methods.

3.2.3. Interpreting the Bitstream. Once the bitstream is accessible to the modern computing environment, document reconstruction can take place. This is where the bitstream—manipulated using the knowledge of formatting techniques—is displayed as the original digital information object. Depending on how well the MKA has accurately and thoroughly captured all of the storage and formatting techniques, the reconstructed document can be an exact representation of the original document.

3.2.4. Output. The result of going through each of the stages of the DRS would result in a recovered digital information object. Given the variety of file formats, the reconstructed object could be an encapsulated document containing metadata or a simple ASCII-text file. This flexibility gives the DRS the hardness to be a long-term solution.

4. Research Question.

The research question to be explored in this Delphi Study is, “Is the DRS model a potentially useful method for maintaining long-term access to digital documents?” The following sub-questions were developed to answer the research question.

1. What are the strengths of the Digital Rosetta Stone Model?
2. What are the areas in the Digital Rosetta Stone Model that need improvement?
3. What is missing from the Digital Rosetta Stone Model?
4. How does the Digital Rosetta Stone Model compare with other models in relation to maintaining long-term access to digital documents?
5. What are the underlying assumptions of the Digital Rosetta Stone Model?
6. What steps are necessary to begin implementation of the Digital Rosetta Stone Model?
7. Who should undertake development and implementation of the Digital Rosetta Stone? And why?
8. Do the experts have anything else to contribute that does not fit in the previous questions?

5. Methodology

The methodology for this research was a Delphi Study. It involves a group of experts who provide their opinion and insight on a certain topic, thus it uses non-probability sampling. The ideas generated are then analyzed and condensed to determine a level of consensus. The Delphi Technique is performed in a series of rounds. This iterative process of rounds and analysis continues until a consensus or stabilization point has been reached. Stabilization indicates that inter-round answers have not changed beyond an appreciable amount. The opinions are typically annotated using a Likert-type scale ranging from Strongly Disagree to Strongly Agree and for Very Important to Not Important [8] [9]. Because this research is inductive in nature, group consensus will not be the only measure of “success”. The idea generation in and of itself will also be useful to the DRS—the ideas submitted by experts in the field can provide important insights into the strengths and weaknesses of the model.

5.1. The Study Population

The population of interest in this study was the group of people called digital archivist experts (DAEs), whose knowledge about the subject area is key to exploring the potential of the DRS Model. Those who constitute the digital archivist community include Information Technology (IT) specialists who are responsible for maintaining long-term access to digital information and are primarily librarians, digital archivists, and academicians. Individuals in this community may be found in a wide variety of industries and government agencies. Key technology makers were considered because of their impact on technology. The organizations under consideration were asked to decide who was most suited to participate in this study.

5.2. The Participants

The organizations that agreed to participate in the study are shown in table 2.

<table>
<thead>
<tr>
<th>Table 2: Participants in the Delphi Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bellcore</td>
</tr>
<tr>
<td>Connectex</td>
</tr>
<tr>
<td>The National Archives and Records Administration</td>
</tr>
<tr>
<td>The RAND Corporation</td>
</tr>
<tr>
<td>The Syracuse University Library</td>
</tr>
<tr>
<td>The University of Pittsburgh, School of Information Science</td>
</tr>
<tr>
<td>The United States Air Force Historical Research Agency</td>
</tr>
<tr>
<td>WGBH</td>
</tr>
</tbody>
</table>

Some experts worked in groups in their respective organizations to develop the answers. The result is that as many as 12 to 15 people actually contributed to this study. For purposes of the technique, the individuals in the group knew who was participating but did not know who made which comments – anonymity has been shown to increase creativity and idea generation [9].

5.3. Implementation of the Delphi Group Study

A description of the DRS Model was sent out to each member of the group, along with instructions on how to participate in the study. The goal of this round was to generate as many ideas about the DRS as the experts felt appropriate. These ideas formed the basis for beginning to develop a consensus.

The initial round took five weeks, after which the results were collated and organized for further group consideration in the second round. In the second round, the members of the Delphi group were asked to agree or disagree, on a five-point Likert Scale with the condensed results and to refine their statements regarding the DRS Model. In this round they were asked to address the issues in terms of both levels of agreement and levels of importance of each of the items, using a 5 point Likert scale.

One of the goals with a Delphi study is to strive for consensus among the group members on the various issues under consideration. Consensus is a measure of how much people agree with one another. For this study, consensus for agreement and high importance was defined as a median of one-half a point above the middle of the Likert scale or higher. Consensus for disagreement or low importance was defined as a median of one-half a point below the middle of the Likert scale or lower. A group consensus that was in the middle of the Likert scale resulted in an unsure rating, for either importance or agreement. There was no distinction between a group consensus of unsure and non-consensus. Non-consensus was for all intents and purposes defined as a consensus of unsure.
The third round was based on the results received from the second round, and allowed the group to continue to refine its position and to work toward consensus. The group members were asked to comment on the report from the second round and to assess it for accuracy and completeness as well as to state their overall assessment of it. The response time for this round was about one week.

6. Results and Analysis

In the initial round, participants were asked to answer questions about the DRS. These results were analyzed using content analysis. The resulting assessment from the first round statements formed the basis for the rest of the study. The statements were categorized into the following eight areas or topics.

1. Preservation and Access Environment
2. Media and Digital Objects
3. Development of the DRS
4. DRS Focus
5. DRS Methodology
6. Metaknowledge Archive
7. Software, Logical Formats and Physical Formats
8. DRS Implementation Details

The response rate for the first round is as follows:
No response received…………………….2 (22.2%)
Responses received………………………7 (77.8%)
Number of total statements received……..66
Number of unique statements received…..54

The opinions expressed during the second round were predominantly recorded in the form of Likert-type scales. At the end of each question, however, there was room for additional comments.

The experts submitted opinions about the issues in two ways. The first opinion related to whether or not the expert agreed with the statement. The second opinion dealt with whether or not the statement was important to the DRS. The opinions were recorded using a 5-point Likert-type scale, with the low end being either disagree or not important. Thus, each of the statements has two opinion parts: statement agreement and statement importance, with three options for each. This results in nine possible responses for each statement.

Table 4: Categories for Opinions

Not every one of the nine categories for opinions had every statement topic in it, but all of the topics fit into the categories. The responses to the second round were analyzed and organized into a report that was then returned to the Delphi group participants for the next round.

The third round consisted of sending the Second Round Report to the group. The group was requested to review the report and comment on any portion of the report that it felt was appropriate. They were asked to see if the generalizations made sense and were reasonable assessments of the second round opinions. Seven of the nine remaining group members responded to this round. The round three responses indicated a high approval of the round two report.

There were three participants who had originally agreed to participate but did not take part in any round. They were not included in the response rates. Of all of the participants, four participated in every round. Five took part in two rounds. No one participated in just one round. Overall, nine experts participated in this study at one point or another.

6.1. Research Questions Answered

Based on high levels of agreement of statements submitted by the Delphi Group, the following answers were provided to the research questions.

1: What are the strengths of the Digital Rosetta Stone Model?
   - It recognizes the importance of retaining access to objects even as the technology for storing them becomes obsolete.
• It allows for access even if no readers for such a medium exists.
• It has the idea of a central registration of document types and specifications.

2: What are the areas in the Digital Rosetta Stone Model that need improvement?
• Where possible, the DRS should integrate well with archiving.
• It does not describe how to handle media degradation and media failure.
• The Metaknowledge criteria needs to be further developed.
• The DRS should place an equal emphasis on the behavior of the software during interpretation of the bitstream and the retrieval process from the physical medium.

3: What is missing from the Digital Rosetta Stone Model?
• The awareness of other long-term access efforts and its compatibility with them.
• The need for self-describing media.
• It does not address the problem of authenticity, or integrity, of the original document.
• It does not address verification and validation of the translation.
• It misses the importance that software plays in interpreting the digital documents by the fact that the behavior of such software is not implicit in a digital artifact's format.

4: How does the Digital Rosetta Stone Model compare with other models in relation to maintaining long-term access to digital documents?
• Other schemas are geared toward digital document preservation.

5: What are the underlying assumptions of the Digital Rosetta Stone Model? If the DRS makes the assumption, is it valid?
• All assumptions identified for the DRS were considered valid.
• The DRS assumes we are in a situation that needs digital archaeology.
• The “native format” is what the original application created.
• Some preserved digital documents will be textual.
• Cooperation with the public and private sectors is necessary.

6: What steps are necessary to begin implementation of the Digital Rosetta Stone Model?
• Clarify whether the model depends on the original medium being available at the time of need.
• Assuming we are ready for a decision, clarify how the model would attempt to recover non-textual information.
• Assuming a feasibility study has been performed, consider the total life cycle costs and probability of the model being successfully implemented.
• Development of the consortium to further build the model.
• The DRS warrants significant investigation at this time.

7: Who should undertake development and implementation of the Digital Rosetta Stone? And why?
• A consortium of those who use and store information.

8: Do the experts have anything else to contribute that does not fit in the previous questions?
• This project needs to be brought into the contact of others where substantive work in this field is being done.

7. Discussion and Recommendations

This research represents an initial assessment of the Digital Rosetta Stone Model by the expert community. As recommended by Robertson [19], it presented the model to the archival community and other interested parties. Some of their overall statements in the first round suggested that the overall impression of the DRS was negative. They addressed problems with the practicality of such an undertaking and that the DRS’s focus may be misguided. However, when asked to address the research questions, their answers proved to be realistic but hopeful.

The group agreed that the DRS recognizes the importance of retaining access to digitally stored documents, and that the DRS could be a possible solution to at least part of the long-term access issue. They noted that many models deal with preservation and associated methods to ensure bit survival. The DRS assumes preservation of the bitstream on the storage medium, and focuses on how to recover and interpret that bitstream. The believed that the preservation and access community should continue to explore the concepts of the DRS with existing and potential preservation strategies in mind. Symbiosis
between these preservation strategies and the DRS could then be nurtured.

The group also recognized the importance of having the right information in the MKA. Therefore, careful and thorough development of the MKA would be a necessary step for a successful DRS. Perhaps hardware and software engineers can uncover this metaknowledge through the research of old technical and scientific journals, as well as the U.S. Patent and Trademark Office or standards-based groups. Among the information that should be stored in the MKA, according to the group, would be knowledge of how best to handle and store the media until bitstream retrieval has occurred. Oftentimes, ignorance in the handling of sensitive objects can undo years of preservation. This is one area where working with preservation specialists could reduce the amount of media failure and slow the process of media degradation caused by abuse or neglect.

In developing the DRS, the experiences gained from others who have worked on access strategies needs to be taken into account. This could help overcome unforeseen obstacles in DRS development or help to build a more robust DRS.

The group suggested that self-describing media could help fill in for the MKA if there are any gaps or inconsistencies and could reduce reliance on a concept such as the DRS. This is where the medium itself has written instructions on how to recover the information. They also raised the issue that when reconstructing the original document, unless there is a human-readable copy or other known stored instance of it, verifying that the output is exactly the same as when viewed using the “native software” will be difficult. Of course, if one already has a readable copy of the document, the need for reconstructing it from the bitstream seems to be moot.

Efforts to develop the DRS and MKA will be expensive and time-intensive. A consortium needs to be developed to conduct extensive research in order to build a robust solution to the long-term access problem. It is therefore necessary to know what resources will be required and when, to best manage development and implementation. Because the DRS would be the link to our obsolete digital history, once developed, it needs to undergo significant testing, and constant upkeep.

Involving a consortium of those who use and store information will benefit the development and implementation of the DRS. As previously mentioned, pitfalls can be avoided and a well-designed solution crafted by those who have worked on previous access or preservation strategies, as well as others who are the beneficiaries of the DRS process. Because the framework needs to be further developed, the DRS project needs to be brought into the contact of those, as mentioned above, to clarify and develop the DRS. If the DRS is developed in isolation, it may not be as comprehensive as it needs to be.

8. Conclusions

The results of the Delphi study of the DRS suggests that it or something like it warrants further study as a means to address the important issue of maintaining access to the increasing volume of digital documents that we continue to create, and which we know are at risk due to technological obsolescence.

9. Limitations

This study has a number of limitations. For instance, the use of the Delphi Technique does not guarantee truth. It works toward group consensus, however, expert groups are not always right. It may be discovered, in time, that other new technological solutions and/or standards better mitigate the risks that stranded data face.

As with Robertson’s study [19], this research does not test the technological feasibility of the DRS. It may turn out that other strategies may be more cost effective, although not necessarily providing a level of assurance that the DRS provides. However, until someone attempts to develop the MKA and build a prototype of the model, the cost aspect of building and using the DRS may not be fully appreciated.

10. Recommendations for Future Research

The group has agreed that the DRS warrants further investigation, and it needs to be brought into the contact of others where substantive work in this field is being done. An important next step may be to design and build a prototype of the DRS and demonstrate its technological feasibility with the help of software and hardware technologists. Showing its practicable efficacy could then lead to a full-scale development and implementation of the DRS.

11. References


