

Digital Preservation and BBC Domesday

Paul Wheatley

University of Leeds

p.r.wheatley@leeds.ac.uk

Presented at the Electronic Media Group

*Annual Meeting of the American Institute for Conservation of Historic and Artistic Works
Portland, Oregon*

June 14, 2004

Abstract

The CAMiLEON Project [1] was initiated in 1999 at the Universities of Leeds (JISC funded) and Michigan (NSF funded) to explore, develop and evaluate technical strategies [2] for use in digital preservation. In particular, CAMiLEON was tasked with getting to grips with a proposed strategy called Emulation. An emulator is a software program that mimics the function of another computer. In a digital preservation context it allows the execution of software programs and data from older computers on a current computer, acting as an alternative to data migration [3]. CAMiLEON developed a range of technical strategies and tested them with practical implementations. This paper examines the work of the University of Leeds in rescuing and emulating an early multimedia resource named BBC Domesday, and its implications for preserving our current digital heritage for the future.

What is BBC Domesday?

In 1983 with the 900th anniversary of the Norman Domesday Book fast approaching, a BBC Television producer named Peter Armstrong came up with an idea for a new digital Domesday that would record detail about life in the United Kingdom in the late 20th century. Rather than resulting in a printed record, BBC Domesday would utilise an exciting new technology called multimedia.

The project resulted in a groundbreaking digital resource that captured the essence of everyday life in the UK in the mid eighties¹. Hundreds of thousands of school children from across the country, a team of 60 researchers at the BBC and countless other scholars, statisticians and photographers contributed to this mammoth resource.

A specially developed software interface allowed the user to search and browse the recorded information that was stored on 2 Videodiscs – the forerunner of today's CDs and DVDs. An adapted BBC Microcomputer², common in schools and homes at the time, was used to view the resource.

Unfortunately, the price of a complete Domesday system was pushed up by 400% as the project went to the market in 1986. At around 4000 pounds sterling a time, the system did not sell well and quickly became a commercial flop and an embarrassment for the

¹ Note that BBC Domesday was a completely new resource about the UK in the 1980s and not a digitisation of the Norman Domesday book.

² The BBC Micro had a similar ubiquity in the UK to that of the Apple II in the USA in the mid eighties.

BBC. Nearly 20 years on and BBC Domesday is viewed in a very different light. It is seen as masterpiece of design and organisation, and a landmark in the development of multimedia that failed only because it was way ahead of its time.

Life and Death

Like most modern multimedia software and data, BBC Domesday was intricately tied to the computer hardware, operating system and software upon which it was based. Consequently the same original hardware setup of an adapted BBC Master Microcomputer and a special Philips LVR0M Videodisc player is needed to use it. Ironically, the 12" videodiscs themselves, upon which BBC Domesday is encoded, have largely survived the last 20 years very well. The original Philips marketing literature describes how the discs were designed to last 100 years. But of course the hardware which reads the discs and runs the software is not so long lasting. This is an important lesson to learn for future. Long lived media does not equal long lived preservation.

CAMiLEON began to look at BBC Domesday when it was already a long way down the gradual slide to digital decay and ultimately oblivion. Figure 1 is an attempt to chart the decline and resurrection of BBC Domesday.

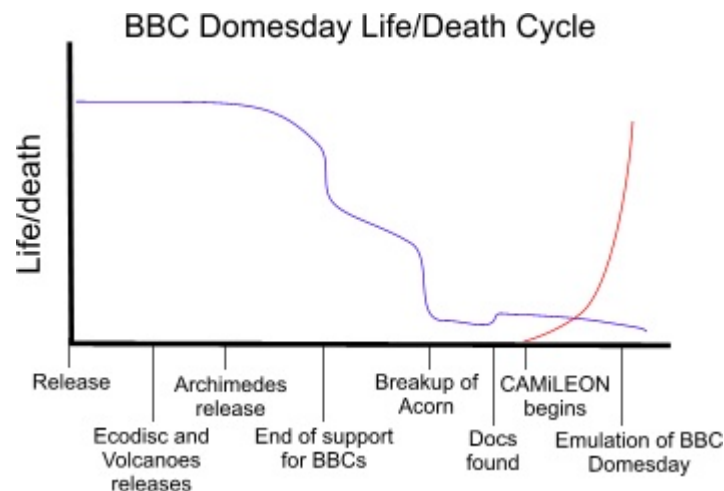


Figure 1

Originally released in 1986, the system quickly became obsolete with the demise of videodisc technology that never achieved large scale uptake outside the U.S.A. The videodisc was replaced by CD, DVD and other rapid advances in computing technology. Two years after Domesday became available, Acorn left the 8 bit BBC Micro platform behind and released the first ever 32 bit RISC computer, the Archimedes. Before long, official support was no longer provided for the BBC and 3rd party support waned as schools across the UK gradually replaced their large collections of BBC Micros (this was of course several years before the dominance of the Windows PC). Years later, financial difficulties led to the break up and demise of Acorn, the technology division being sold off to Pace Microelectronics. Discovery of some Domesday documentation at

the BBC which now resides in the BBC Archives at Reading led to a slight recovery in the gradual life/death decline - much of the project documentation was lost or scattered when BBC Enterprises ceased to exist. The new upwards curve represents the work of CAMiLEON and others in reviving and preserving BBC Domesday.

Looking to the past to inform the future

CAMiLEON was not tasked with rescuing digital resources in danger of obsolescence but it was felt that practical work should both inform the design of preservation strategies and act as a test of how effective those strategies would work in a real situation. A rescue of a 1970s mainframe named George 3 [4], in conjunction with the Computer Conservation Society [5] was undertaken at the start of the project and fed into the design of CAMiLEON's emulation strategy. In the final year of the project, a rescue of BBC Domesday and development of a working emulation demonstrator was undertaken.

CAMiLEON identified that work to preserve current materials was very hard to evaluate. A PDF file, when double clicked, will automatically be launched and displayed on a modern computer by Windows and some Adobe software. Illustrating the ability to render a PDF at some point in the future when Windows no longer runs on a current computer is hard to simulate. However, by examining digital resources from the 1970s or 1980s and seeing how they could have been preserved to maintain the ability to render them over time is eminently possible and is a far more realistic test of digital preservation strategy.

Why tackle Domesday?

The key reason for taking on BBC Domesday as a proof of concept emulation demonstrator was the sheer complexity of the task that was faced. The project team quickly found that BBC Domesday acted as a perfect stimulus for debating digital preservation processes. At the end of the day, if rescue and emulation of BBC Domesday could be achieved, the potential of emulation in digital preservation would become apparent. A number of other factors also made Domesday a good choice:

- It is a complex interactive object that would be difficult to preserve in its entirety using traditional preservation strategies (eg. migration)
- It relies on a number of peripheral devices which was an area which deserved further exploration
- Several other resources were produced after BBC Domesday using the same technology. CAMiLEON hoped to show that preservation work of this kind could be cost effective when applied to a number of similar digital objects
- It represents a valuable historical and intellectual resource in grave danger of loss without urgent preservation work.

- BBC Domesday is frequently used as an example of the digital preservation problem and hence its preservation could act as an exemplar for further work in this field

Rescue and preservation

The term “digital archaeology” is sometimes used to describe the process of rescuing a digital resource that has become obsolete. The work undertaken by CAMiLEON in rescuing BBC Domesday could certainly be categorized in this way. The very first stage of the work was to get hold of a BBC Domesday system and discover what it looked like, how it worked, and how it could be rescued.

The team was fortunate enough to take custody of a broken Domesday setup from the University of Leeds’ Geography department. A process of discovery and electronic repair began. The act of opening up the BBC Master for the first time and discovering what lay in wait resulted in the discovery of some retro fitted boards that turned out to be a co-processor board and a SCSI board. Both had never been emulated before.

Repair work by the team got the kit working and a range of research activities began to turn up valuable contacts of technical personnel who developed Domesday in the eighties and some documentation which would prove invaluable. The discovery of the technical manual for the LVRom videodisc player was a considerable step forward. After debate amongst the team, work on rescuing the data and emulating the computer platform began in earnest.

Getting the bits

Utilising and testing techniques developed on CAMiLEON’s sister project, CEDARS [6], the project team developed through a process of trial and error a mechanism for extracting the data from videodiscs. The data would be held in a media neutral form which could easily be copied from storage device to storage device to facilitate its preservation on into the future. The team designed an appropriate Underlying Abstract Form [7] that would adequately capture all the properties of the original that were deemed to be significant.

The complex and sizeable nature of the resource provided a particularly good test of the implementation of the UAF concept.

BBC Domesday utilised videodiscs as its storage medium, in the years before Compact Disc technology became available. Data is recorded on the BBC Domesday videodiscs in two different forms. Due to the limitations of computer graphics hardware at the time, image data was stored in a similar form to standard videodisc movies. A still image format was devised from the standard movie format. An analogue video signal for each still image is directly frequency modulated into sub carriers. The resulting pulse-width-

modulated waveform is used to derive the pits and troughs for recording on to the disc³. After considerable debate amongst the project team it was decided that the encoding information and waveform was not significant. Only the final images in terms of form, shape and colour, and the frame number of each image was significant. The only form of indexing used by the Domesday software with regard to the image data, was based on the frame number of each image. Consequently a process of “semi-digitisation” was performed to capture the images in a suitable form for preservation. The images were digitised at a resolution equivalent to the number of vertical lines in the TV signal output and at a horizontal resolution that resulted in a sensible aspect ratio.

The “digital” data present on the discs (composing a range of textual, statistical and mapping data) was encoded using a similar technology to the image data and stored on the videodiscs in the space normally used to store the sound track for a standard videodisc. Again it was decided that the encoding and waveform data was not significant. The digital data was grabbed on a sector by sector and block by block basis and mapped to a single bytestream that was composed as a disc ID header followed by a sequence of data blocks. These blocks are interpreted by the original Domesday system as a Video Filing System (VFS) similar to the then standard BBC Disc Filing System (DFS).

Because of the unusual design of BBC Domesday and the “semi-digital” storage used to record the image data, the process of transforming the resource from hand held media to an easily preserved Underlying Abstract Form was a complex one. The pitfalls of not considering this process in detail are very apparent. A simple copying of the contents of the “digital” data on the discs would not have captured the disc ID header, without which the original software cannot run (leaving preservation by emulation a difficult route).

Emulation

The project team worked from the basis of an existing open source BBC Micro emulator and developed the software to provide support for the additional hardware elements required to run BBC Domesday. These included support for the extra facilities of the BBC Master, the SCSI hardware, the co-processor, and of course the LVR0M Videodisc player.

The LVR0M VP415 was designed specifically for use with BBC Domesday. The LVR0M communicated with the BBC Master via a SCSI interface which was retro fitted to the BBC Master for use in Domesday. The LVR0M performs a complex role of both serving digital data to the BBC as well as generating an RGB signal of still videodisc images that is synced and mixed with the TV signal from the BBC Master, providing different modes of overlay and transparency. This unusual setup required a substantial amount of hardware functionality to be emulated in software. The digital data was represented simply as a bytestream for each side of the discs (as described above). Emulation of the

³In effect, the storage is a digital signal with variable (analogue) timing.

LVRom video functions were integrated into the BBC Master emulator due to the requirements for video mixing. The LV DOS command set was chosen as the level of abstraction to be emulated given its simplicity, ease of integration with the BBC Master emulator and availability of documentation.

Progress in developing the emulator was not easy, particularly in emulating the co-processor due to the lack of documentation. The team was however successful and the result was a working emulator which ran BBC Domesday without any of the hardware from the original setup.

Software Longevity

CAMiLEON invested considerable effort in developing techniques to ensure emulator software would not itself become obsolete due to encroaching technology obsolescence. The key to successful longevity is to identify appropriate levels of abstraction both in the target platform to be preserved and the current platforms upon which preservation software will execute. CAMiLEON noted that data formats, computer hardware and computer operating systems had very limited longevity (even file format *standards* saw little longevity). But programming languages offered a lot more stability and longevity. CAMiLEON worked to develop preservation subsets that would more easily facilitate software longevity [4]. Follow up research and development on the “Representation and Rendering Project” [8] explored the effective use of JAVA in migration tools that could be maintained and development will into the future.

Timeliness

As the rescue of BBC Domesday highlights, the timeliness of preservation action is crucial. If action is left too long, the effort required to preserve is greatly increased. Had the rescue of Domesday been undertaken at the time of the demise of the BBC Micro platform, the cost of the action and the level of complexity of the work involved would have been much reduced. As figure 1 shows however, the process of digital obsolescence is often a gradual one. This illustrates the need for technology watch processes that will monitor current technologies and provide alerts as to when and where preservation action needs to be taken.

CAMiLEON attempted (in some cases in conjunction with the Computer Conservation Society) the rescue of digital resources from 3 different decades. Rescues from the 70s and 80s were possible but were complicated by the lack of working systems, documentation and personal experience of the original systems. The rescue of materials from the dawn of computing, in this case one of the Leo computers, had to be abandoned despite the discovery of some vital source code. The conclusion was simply that not enough was known about the design of the machine to enable it to be resurrected using emulation.

The bigger picture

Whether preserving a simple digital object or an example as complex as BBC Domesday, a digital repository will require far more than just the rendering mechanism (the emulator in the case of Domesday) in order to ensure effective preservation. OAIS [9] describes the technical metadata used to record the structure and semantics of a preserved object as Representation Information.

Work to define how this metadata might be organised and what mechanisms would facilitate its use, management and monitoring is at an early stage, but the digital preservation team at Leeds has conducted a range of work in this area. Beginning with the Cedars project, CAMiLEON, and continued most recently as part of the Representation and Rendering Project, the team has developed a practical framework for the implementation of a Network of Representation Information [10]. Part of this JISC funded research and development included the construction of a Representation Information Case Study. This Representation Network (a navigable collection of related technical metadata) details the mechanisms used to rescue the BBC Domesday data, and illustrates some of the technical metadata that a digital repository might record about such an object.

Recording sufficient technical metadata to enable effective long term preservation within the constraints of realistic levels of funding is a pressing problem for the preservation community. Gradually evolving networks of Representation Information that can be globally shared, tested, updated and re-used is a possible solution that is discussed in more detail by Holdsworth and Wheatley [11].

Conclusions

At the time of the launch of the CAMiLEON Project back in 1999, there were some proponents of emulation, in particular Jeff Rothenberg [12]. But on the whole, emulation was viewed with a sceptical eye. Many thought that at best it would be prohibitively costly and at worst it simply would not work. CAMiLEON has contributed to a gradual change in perception and most now recognise that emulation does at least have some part to play.

Obviously the model of developing a specific emulator in order to preserve a single digital resource is not a realistically scalable one. But of course, the beauty of emulation is that many digital objects of different type and format will all run on the same emulator. A number of different computer platforms will need to be emulated to provide a useful coverage of support but there are far fewer computing platforms than there are data formats. Used carefully emulation could even be cheaper than migration.

Using the right techniques to keep costs down will also be crucial. Software longevity and the use of high level abstractions, where possible, are absolutely essential. Fore thought must be given to making a particular emulation survive beyond its current implementation without undue expenditure of effort. A process of abstracting out low

level technical operation and emulating higher level function, leads to more meaningful and simpler preservation. This is particularly relevant with peripheral devices, many of which are similar in function but very different in implementation.

Some hurdles certainly remain, particularly in the realm of documentation and user metadata. If we preserve both the data and the original access method of a particular digital object, users of the future need to know how the unfamiliar user interface from the past works. Much more practical testing and experimentation is needed before emulation can become a practical tool for general users of the future. But even at this early stage, emulation is already a viable tool for the digital preservationist. A key problem with migration is the very real danger of losses in accuracy as a number of migrations are performed over time. Verifying that a migrated object is accurate is very difficult when the original has long been lost to technology obsolescence. Emulation can provide that comparison and verification.

The power of BBC Domesday to capture the public imagination has yet to die, nearly 20 years on from its initial launch. Various announcements related to the work of the CAMiLEON project and the launch of the Digital Preservation Coalition in the UK attracted a great deal of media interest which JISC has worked with to raise the profile of digital preservation work within the public arena [13]. While certainly not a core aim of the project, this is a significant achievement of the work.

Where next for BBC Domesday?

Unfortunately CAMiLEON was not able to make the fruits of its rescue work available to the public due to the complex copyright situation surrounding Domesday. However, some progress has been made in this area and a team at the University of Leeds are currently seeking funding to turn the CAMiLEON demonstrator into a system capable of public access to the majority of the Domesday resource.

The work on the emulation of BBC Domesday has been short listed for the 2004 Digital Preservation Award [14].

References

- [1] "CAMiLEON", <http://www.si.umich.edu/CAMiLEON/>
- [2] "PADI - Preservation Strategies" <http://www.nla.gov.au/padi/topics/18.html>
- [3] "PADI - Data Migration", <http://www.nla.gov.au/padi/topics/21.html>
- [4] "C-ing ahead for digital longevity", Holdsworth, D, <http://www.si.umich.edu/CAMiLEON/reports/cingahd.html>
- [5] "Computer Conservation Society", <http://www.cs.man.ac.uk/CCS/>

- [6] "CEDARS", <http://www.leeds.ac.uk/cedars>
- [7] "CEDARS Guide to Technical Strategies"
<http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html>
- [8] "Representation and Rendering Project : Representation Networks",
<http://www.leeds.ac.uk/reprend/repnet/repnet.html>
- [9] Reference Model for an Open Archival Information System (OAIS) ISO 14721:2002,
<http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>
- [10] "A blueprint for Representation Information in the OAIS model", Holdsworth, D and Sergeant, D, <http://esdis-it.gsfc.nasa.gov/MSST/conf2000/PAPERS/D02PA.PDF>
- [11] "Long term stewardship of globally distributed representation information"
Holdsworth, D and Wheatley, P,
<http://romulus.gsfc.nasa.gov/msst/conf2004/Papers/MSST2004-03-Holdsworth-a.pdf>
- [12] "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation", Rothenberg, J,
<http://www.clir.org/pubs/reports/rothenberg/contents.html>
- [13] "BBC Domesday Publicity",
<http://www.si.umich.edu/CAMILEON/domesday/press.html>
- [14] "2004 Digital Preservation Award",
<http://www.dpconline.org/graphics/awards/shortlist.html>