

DOI:10.1145/2160718.2160728

David Anderson

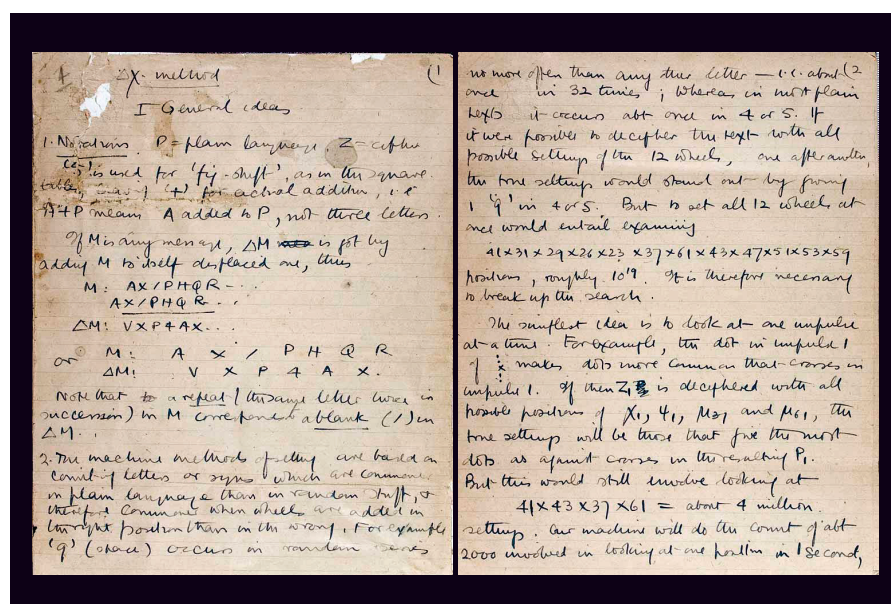
Historical Reflections

The Future of the Past

Reflections on the changing face of the history of computing.

I HAVE ALWAYS thought that Erasmus had his priorities correct when he remarked: “When I get a little money, I buy books; and if any is left I buy food and clothes.” For as long as I can recall I have loved secondhand bookshops, libraries, and archives: their smell, their restful atmosphere, the ever-present promise of discovery, and the deep intoxication produced by having the accumulated knowledge of the world literally at my fingertips. So it was perhaps inevitable that, despite having pursued a career in computer science, I would be drawn inexorably toward those aspects of the discipline that touched most directly on the past: the history of computing and preservation.

A great deal of my personal research, particularly on the largely unknown role played by M.H.A. Newman in the post-war development of computers, has obliged me to spend many hours uncovering and studying old letters, notebooks, and other paper-based records. While some of my source material came to me after having been transferred to digital format, none of it was born digital. Future historians of computing will have a very different experience. Doubtless they, like us, will continue to privilege primary sources over secondary, and perhaps written sources will still be preferred to other forms of historical record, but for the first time since the emergence of writing systems some 4,000 years ago, scholars will be increasingly unable to access directly historical material. During the late 20th and early 21st century, letter writing has given way to email, SMS messages, and tweets,



Cracking German codes during WWII. Handwritten notes by M.H.A. Newman.

diaries have been superseded by blogs (private and public), and where paper once prevailed digital forms are making inroads and the trend is set to continue. Personal archiving is increasingly outsourced—typically taking the form of placing material on some Web-based location in the erroneous belief that merely being online assures preservation. Someone whose work is being done today is likely to leave behind very little that is not digital, and being digital changes everything.

Digital objects, unlike their physical counterparts, are not capable directly of human creation or subsequent access but require one or more intermediate layers of facilitating technology. In part, this technology comprises further digital objects: software such as a BIOS, an operating system, or a word

processing package; and in part it is mechanical, a computer. Even a relatively simple digital object such as a text file (ASCII format) has a surprisingly complex series of relationships with other digital and physical objects from which it is difficult to isolate it completely. This complexity and necessity for technological mediation exists not only at the time when a digital object is created but is present on each occasion when it is edited, viewed, preserved, or interacted with in any way. Furthermore, the situation is far from static as each interaction with a file may bring it into contact with new digital objects (a different editor, for example) or a new physical technology.

The preservation of, and subsequent access to, digital material involves a great deal more than the safe

storage of bits. The need for accompanying metadata, without which the bits make no sense, is understood well in principle and the tools we have developed are reasonably reliable in the short term, at least for simple digital objects, but have not kept pace with the increasingly complex nature of interactive and distributed artifacts. The full impact of the lacunae will not be completely apparent until the hardware platforms on which digital material was originally produced and rendered become obsolete, leaving no direct way back to the content.

Migration

Within the digital preservation community the main approaches usually espoused are migration and emulation. The focus of migration is the digital object itself, and the process of migration involves changing the format of old files so they can be accessed on new hardware (or software) platforms. Thus, armed with a suitable file-conversion program it is relatively trivial (or so the argument goes) to read a WordPerfect document originally produced on a Data General minicomputer some 30 years ago on an iPad 2. The story is, however, a little more complicated in practice. There are something in excess of 6,000 known computer file formats, with more being produced all the time, so the introduction of each new hardware platform creates a potential need to develop afresh thousands of individual file-format converters in order to get access to old digital material. Many of these will not be produced for lack of interest among those with the technical knowledge to develop them, and not all of the tools that are created will work perfectly. It is fiendishly difficult to render with complete fidelity every aspect of a digital object on a new hardware platform. Common errors include variations in color mapping, fonts, and precise pagination. Over a relatively short time, errors accumulate, are compounded, and significantly erode our ability to access old digital material or to form reliable historical judgments based on the material we can access. The cost of storing multiple versions of files (at least in a corporate environment) means we cannot always rely on being able to retrieve a copy of the original bits.

The challenges associated with converting a WordPerfect document are simpler than those of format-shifting a digital object as complex as a modern computer game, or the special effects files produced for a Hollywood blockbuster. This fundamental task is well beyond the technical capability or financial wherewithal of any library or archive. While it is by no means apparent from much of the literature in the field, it is nevertheless true that in an ever-increasing number of cases, migration is no longer a viable preservation approach.

Emulation

Emulation substantially disregards the digital object, and concentrates its attention on the environment. The idea here is to produce a program that when run in one environment, mimics another. There are distinct advantages to this approach: it avoids altogether the problems of file format inflation, and complexity. Thus, if we have at our disposal, for example, a perfectly functioning IBM System/360 emulator, all the files that ran on the original hardware should run without modification on the emulator. Emulate the Sony PlayStation 3, and all of the complex games that run on it should be available without modification—the bits need only be preserved intact, and that is something we know perfectly well how to accomplish.

Unfortunately, producing perfect, or nearly perfect emulators, even for relatively unsophisticated hardware platforms is not trivial. Doing so involves not only implementing the documented characteristics of a platform but also its undocumented features. This requires a level of knowledge well beyond the average and, ideally, ongoing access to at least one instance of a working original against which performance can be measured.

Over and above all of this, it is critically important to document for each digital object being preserved for future access the complete set of hardware and software dependencies it has and which must be present (or emulated) in order for it to run (see TOTEM; <http://www.keep-totem.co.uk/>). Even if all of this can be accomplished, the fact remains that emulators are themselves software objects written to run on par-

ticular hardware platforms, and when those platforms are no longer available they must either be migrated or written anew. The EC-funded KEEP project (see <http://www.keep-project.eu>) has recently investigated the possibility of developing a highly portable virtual machine onto which emulators can be placed and which aims to permit rapid emulator migration when required. It is too soon to say how effective this approach will prove, but KEEP is a project that runs against the general trend of funded research in preservation in concentrating on emulation as a preservation approach and complex digital objects as its domain.

Conclusion

Even in a best-case scenario, future historians, whether of computing or anything else, working on the period in which we now live will require a set of technical skills and tools quite unlike anything they have hitherto possessed. The vast majority of source material available to them will no longer be in a technologically independent form but will be digital. Even if they are fortunate enough to have a substantial number of apparently well-preserved files, it is entirely possible that the material will have suffered significant damage to its intellectual coherence and meaning as the result of having been migrated from one hardware platform to another. Worse still, digital objects might be left completely inaccessible due to either not having a suitable available hardware platform on which to render them, or rich enough accompanying metadata to make it possible to negotiate the complex hardware and software dependencies required.

It is a commonplace to observe ruefully on the quantity of digital information currently being produced. Unless we begin to seriously address the issue of future accessibility of stored digital objects, and take the appropriate steps to safeguard meaningfully our digital heritage, future generations may have a much more significant cause for complaint. □

David Anderson (cdpa@btinternet.com) is the CiTECH Research Centre Director at the School of Creative Technologies, University of Portsmouth, U.K.

Copyright held by author.