The Digital Dark Age, and why it will have to wait.

The growth in accessible digitized primary source materials, together with the improvement in in quantity and quality of online research tools has changed fundamentally the way in which historians go about their business. While the fundamental importance of primary sources remains the same, it is ever more the case that the sources themselves were born digital, consumed digitally, and (hopefully) preserved digitally. But what assurance do we have that in the next century, or the next millennium, historians will be able to access materials which were produced on machines, and software packages which have long ceased to exist? This is the question that Google Vice President and Chief Internet Evangelist, Vint Cerf has raised in a number of engaging recent public talks and interviews. His conclusion is that we are standing on the edge of a precipice, and that unless we take appropriate steps a digital dark age awaits us. I cannot easily gauge the effect of Cerf's remarks in the USA, but on this side of the pond, his comments have been taken up widely, and accepted more or less uncritically, by the broadcast and print media. Among professionals working in the Digital Preservation field, the reaction has been much less accepting. Many of the people with whom I come into contact exhibit a simmering resentment that a great deal of pioneering and fundamental work carried out in Europe (and more widely), is being overlooked in the publicity storm that Cerf's pronouncements have generated.

This is entirely understandable, but I think that it is almost certainly the wrong response. For most of the last decade, I have been working actively in digital preservation, having been drawn into the field from a background in what has come to be called the "Digital Humanities", where I continue to take a particular interest in the history of Computing. For most of that time, it has been something of an uphill struggle to persuade academic colleagues that digital preservation was not only something that needed to be taken seriously, but was a field in which there were, and are, interesting and intellectually satisfying challenges to be addressed. What was true of fellow academics was doubly so of business leaders, politicians and other opinion formers. Fellow digital preservationists were often inclined to scratch their heads, and with their thousand-yard stare firmly in place, ask what they have to do to get preservation onto the political and business agenda. I have attended a number of conferences where participants debated whether the best approach to getting preservation taken seriously was to deploy somewhat apocalyptic morality tales of the dangers of losing large slices of our cultural heritage as a result of inattention to preservation, or to encourage better digital custodianship by more positive means. It probably needs a little of both, but the uncomfortable truth is that the messenger is often more important than the message, and whether the approach is to scare the world at large into taking preservation seriously, or to provide less dramatic encouragement, the messenger needs to be able to cut through the noise and be heard. Vint Cerf, as one of the architects of the Internet, and a key figure on the American politico-scientific stage, is extremely well placed to bring digital preservation to greater prominence than has hitherto been the case, and should therefore be welcomed wholeheartedly, even if he does not always capture fully the digital preservation zeitgeist. The media interest in Cerf's talk of a looming digital dark age can only serve to help raise awareness among computer scientists and engineers and encourage them to engage with the outstanding technological challenges. His intervention may even be instrumental in persuading funders to put their dollars and euros behind much-needed digital preservation research.

The almost daily news of the damage being caused to historic buildings, and sites of archaeological and cultural importance in the Syrian Civil war, makes talk of a digital dark age seem particularly resonant. The recent execution by Isis of the long-time keeper of Palmyra's extraordinary cultural artefacts, octogenarian Khaled al-Asaad, for 'crimes' including representing Syria at "infidel conferences", and serving as "the director of idolatry" in Palmyra, should remind us that in some parts of the world, protecting cultural heritage comes at a very high price. However, whatever grounds there are for despondency in the Middle East, and somewhat at odds with the picture being drawn by Vint Cerf, the situation with the protection of our digital heritage is far from bleak, and there are, indeed, many reasons to be cheerful. Over recent years the global digital preservation community has been very active and there has been real and substantial progress made. In addition to calling for new technologies and techniques, and promoting the "Olive" project at Carnegie Melon (https://olivearchive.org) which approaches preservation by using virtual machines running application state capture files, Cerf has drawn attention to the importance of the human, societal, and organisational dimensions of preservation. His call, for example, to revisit the rules on copyright to allow for a "fair use" provision in the case of preservation activity would be a real step forward.

Despite the impression which is left by Cerf's remarks, is not the only fruit of the digital preservation field. Here in Europe, in response to Cerf, the Executive Director of the Digital Preservation Coalition, Dr. William Kilbride, has issued a call for the preservation community to highlight preservation activities or projects in which they have been involved, so that a fuller picture of the preservation landscape over recent years might be given. Many of these contributions have been gathered together in a twitter feed with the pleasingly optimistic hashtag (https://twitter.com/hashtag/nodigitaldarkage).

For the remainder of this piece, and in the spirit of the nodigitaldarkage campaign, I would like to draw some attention to a number of projects and other activities that members of my own team have led, or in which they have played prominent roles. These cover the development of innovative tools and techniques, dedicated outreach and dissemination, and the organisational aspects of preservation and re-use of culturally significant material in digital form.

Vint Cerf's comments give very little idea of the intellectual roots of projects like Olive . Most of the early work in digital preservation concentrated on migration as a preservation approach. This, in essence, depends on copying or converting digital objects originally intended to run on one technology platform to run on another. Inevitably, migration involves changing some of the characteristics of the original digital object, so a lot of attention must be paid to ensuring that the properties considered to be of the greatest importance for a designated stakeholder community are preserved intact. Major practical limitations in the migration approach are exposed when the digital objects in question are inherently complex. Migrating a modern computer game from one platform to another, for example, involves a level of technical expertise, which simply does not exist in cultural heritage organisations, and involves IPR issues, which are, for all practical purposes insurmountable. As the tendency is for digital objects to become ever more complex, interest has turned to developing approaches to preservation that depend, as Olive does, on emulation. Although not mentioned by Cerf, the KEEP (Keeping Emulation Environments Portable) project (http://www.keep-project.eu), was the first publically funded project to develop emulation services to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc. The overall aim of the project was to facilitate universal access to our cultural heritage by developing flexible tools for accessing and storing a wide range of digital objects. KEEP also considered legal issues concerning the implementation of emulation-based systems and propose solutions which comply with European and national copyright laws.

Digital objects are, of course, connected intimately with the technical environments on which they were created and used. In order to ensure long-term preservation of, and access to, digital material it is therefore essential to record carefully the hardware and software dependencies of each digital object in a preserved corpus. Typical information required includes details of the computer hardware, operating system, plug-ins, software libraries etc., which a preserved object originally required, together with information on the hardware and software environment that was used during any subsequent preservation actions such as migration or emulation. Cerf gives no significant attention to the complexity involved in assembling and maintaining even the basic technical environment metadata required for subsequent emulation. This is a seriously time-consuming, detailed and complex task in its own right. To address this challenge, my colleague Dr. Janet Delve led the development of the TOTEM (Trustworthy Online Technical Environment Metadata) technical registry (http://amzn.to/1JuKR3c). The TOTEM generic data models, a database implementation, and a metadata schema have been combined with a compatible OWL ontology created within the PLANETS project.

Although engaged in public outreach, Cerf does not refer to the considerable efforts that have been made in this area by organisations like the Open Preservation Foundation, or the Digital Preservation Coalition. Nor does he refer to any of the projects which have played a role in this space. The POCOS (Preservation of Complex Objects Symposia) project (http://bit.ly/1LmMaPa) which concentrated not on the development of new tools or techniques, but was established to give global thought-leaders in research into the Preservation of Complex Objects an opportunity to share and thereby extend the body of knowledge on this topic through a series of symposia at locations across the United Kingdom. The fundamental task facing these symposia was to present material of great technological and organisational complexity in a lucid, cogent, relevant and approachable manner so as to engage UK HEI researchers and practitioners in a wide variety of disciplines, as well as reaching those further afield in, for example, commerce, industry, cinema, government, games and films classification boards.

The seminars were arranged around three general themes: visualisations & simulations, software art, and videogames & virtual worlds. Each of these domains involves the development, use, and manipulation of complex digital objects, and each presents a different, although clearly related, set of preservation challenges. A substantial and innovative dissemination programme was established to ensure that the various stakeholder communities obtained the maximum long-term value. This included the production of a peer-reviewed book (http://bit.ly/1NxL3i8) presenting the key outputs.

While it is clear that Cerf recognises the need for integrated approaches to digital preservation that include organisations as well as tools, he does not seem to be aware of the considerable strides which have already been made, and continue to be made in this arena. On the organisational side, archives provide an indispensable component of the digital ecosystem by safeguarding information and enabling access to it. Harmonisation of currently fragmented archival approaches is required to provide the economies of scale necessary for general adoption of end-to-end solutions. There is a critical need for an overarching methodology addressing business and operational issues, and technical solutions for ingest, preservation and re-use.

To address this, the E-ARK project (http://www.eark-project.com/) working co-operatively with commercial systems providers, is creating and piloting a pan-European methodology for electronic document archiving. The emphasis is on not on "blue-sky" research but on synthesising existing tools and techniques which have been developed over the last decade or so both commercially and within the context of publically funded research projects. National and international best practices, that will keep records and databases authentic and usable over time, are also being drawn together and integrated, with the intention of providing a single, scalable, robust approach capable of meeting the needs of diverse organisations, public and private, large and small, and able to support complex data types. E-ARK will therefore demonstrate the potential benefits for public administrations, public agencies, public services, citizens and business by providing simple, efficient access to the workflows for the three main activities of an archive - acquiring, preserving and enabling re-use of information.

The methodology will be implemented in various national contexts, using existing, near-to-market tools, and services developed by the partners. This will allow memory institutions and their clients (public- and private-sector) to assess, in an operational context, the suitability of those state-of-the-art technologies.

The practices developed within the project will reduce the risk of information loss due to unsuitable approaches to keeping and archiving of records. The project will be public facing, providing a fully operational archival service, and access to information for its users. The project results will be generic and scalable in order to build an archival infrastructure across the EU and in environments where different legal systems and records management traditions apply. E-ARK will provide new types of access for business users.

E-ARK will pilot an end-to-end OAIS-compliant e-archival service covering ingest, vendor-neutral archiving, and reuse of structured and unstructured data, thus covering both databases and records, addressing the needs of data subjects, owners and users. The pilot and methodology will also focus on the essential pre-ingest phase of data export and normalisation in source systems. The pilot will integrate tools currently in use in partner organisations, and provide a framework for providers of these and similar tools ensuring compatibility and interoperability. A core component of the project is the integration platform which uses the existing ESSArch Preservation Platform (EPP) application as an Archival Information System, which is already in productive deployment at the National Archives of Norway and Sweden. In order to achieve scalability, E-ARK will adopt a data management and storage layer for this tool on top of the proven open-source Cloudera CDH4 distribution of Apache Hadoop, enabling storage and computational power to be seamlessly added to the system.

All in all, there is considerable reason to feel hopeful that significant progress will continue to be made in digital preservation in the years ahead. Many of the problems that appeared, just a few years ago, to be almost intractable are being brought under control. The pace of technological change continues unabated, and this brings with it, fresh preservation challenges, but despite rumours to the contrary, it appears the digital dark age will have to wait a little while longer.

I would like to dedicate this piece to Khaled al-Asaad.