

INTRODUCTION

When I proposed this topic a few months ago, I mainly intended to talk about the structural changes in hard- and software that will make it more and more difficult to define what a document is and, therefore, what archives should collect in the future. But then, recently, Vint Cerf, one of the ‘fathers of the internet’ and, more specifically, inventors of the internet’s most important protocol TCP/IP and now Vice President and “Chief Internet Evangelist” at Google, gave a talk at the Annual Meeting of the American Association for the Advancement of Science at San Jose. The topic of his talk seemed to anticipate the points I want to make here – and therefore, could make me look like a plagiarist. Fortunately, this is not the case.

At San Jose, Vint Cerf warned of the “bit rot” when he spoke about the obvious problems of digitisation and of the sustainable and future-proof storage of our digital *documents* – be they text documents, e-mails, images, movies or anything else we could call ‘documents’ in the broadest sense. I do not know – yet – of any such solution, and I dare to doubt, that the “digital vellum” Cerf proposed at the conference really would be one. He himself seems to see it – at least, for the moment – this way:

“We are nonchalantly throwing all of our data into what could become an information black hole without realising it. We digitise things because we think we will preserve them, but what we don’t understand is that unless we take other steps, those digital versions may not be any better, and may even be worse, than the artefacts that we digitised,” Cerf told the Guardian. — “If there are photos you really care about, print them out.” [www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf]

Though he also mentioned tweets, the problems connected with this special kind of digital communication and its parallels, to Cerf, seem to be of the same level and quality as for ‘normal’ documents. The “digital vellum”, i.e. the solution Cerf suggested, should consist of some sort of digital emulation of the entire computational infrastructure we use to create our documents: the dedicated software programs like office or e-mail programs, the operating system on which they run in that special version used to create the documents and even the hardware on which the operating system runs. Even if we may leave aside here the problems connected with licensing of all of these parts ‘for eternity’ and the copy- and other rights to use them in such an emulation – and even if this would be possible at all while today some software already protects itself from being run in virtual machines: there is still a problem. Because, in my opinion, such an emulation would only lift all the problems we are facing with the preservation of digitised documents already and in the near future to the next level; one may call it a ‘meta level’. And there, on this ‘meta level’ we – inevitably – will encounter the same problems. Only our ‘documents’ we will have to

keep, to store and reconstruct will be ‘a little bit’ larger, i.e.: entire computational systems *and* their states at any point in time separated from the next by an arbitrary time span. . .

We do not even have to think about some future computer technologies like quantum computers that will not be digital in a narrow sense, e.g. will not calculate with ones and zeros anymore. Surely, such super-computers should be able to simulate a virtual environment for the machines and software from our digital ‘stone age’ without any problems. We may think even more ahead into the future and of systems that would not ‘calculate’ with programs and some sort of bits and bytes but, for instance, rather be comparable to the biological systems like our brain. Even inside of those the simulation of a special hardware, software with its licenses and documents with their admission rights etc. should be possible.

But that is – at least in my opinion – only one, the smaller side of a larger problem.

THE DISAPPEARANCE OF THE DOCUMENT (PARADIGM)

The problem that I would like to point your attention to is already here, and I do not see any suggestion for a solution: While Vint Cerf like many archivists is afraid of the digital future for our documents, i.e. digital files representing them and created with office programs, saved as PDFs, digital images etc. . . documents, for which we do not have any solution that could guarantee their availability for the next 50 or even only 25 years . . . While we are still looking for a solution for *this* problem (how to save our digital documents) – this very document paradigm for digital data is already fading. Of course, also in the foreseeable future we will want to keep pictures (or any other form of digital representation) of our beloved ones or important events, we will want to keep movies from Holly- or Bollywood or films of personal moments and experiences and we will surely (at least I will) want to listen to great performances of music at any place and time . . . and especially: You as archivists will want to keep digital versions of historically important documents of any kind for the longest future possible. We may even suppose that databases could be seen as some sort of very large ‘documents’ that could be saved in Vint Cerf’s “digital vellum”, that we will be able to preserve the constantly changing states of these database files and that even the active work with these databases will still be possible because the software used to manage the database and to interact with the data and to filter information etc. would also be part of Cerf’s “digital vellum”.

But: The digital ‘document’ itself could, can and will be – and even is already – substituted by other forms of communication that do *not* follow the document paradigm – which, in itself, is only a metaphor. – When computers were very large machines with small amounts of memory space and usually handled and administered with cryptic commands, the metaphorical paradigm of a desktop with folders and files containing (text) documents, pictures etc. was invented together with the graphical user interface to make them more

user friendly and to help people to orient themselves through the digital ‘jungle’ hidden in the chips, drives and file systems of these machines.

The aforementioned ‘tweets’ and their software environment *twitter* may serve as an example of what I regard as the coming mainstream of our interaction with computers – leading to the displacement, if not even the replacement of the ‘digital document paradigm’.

Think of a member of the Bundesrat sending an information over twitter like: “The binding of the Swiss Franc to the Euro will be cut tomorrow.” Of course, before this information is printed in newspapers, even before it appears in online magazines, in e-mail newsletters etc., i.e.: before these bits of information are organised or laid down in any form of what could be regarded as a ‘document’ – this one-sentence tweet would be a historical fact and information worth to be documented for the future. But, then, what should the contemporary archivist keep in his (or her) archive? Only this tweet? Only the first 100, 1’000, 10’000 tweet reacting on it? If we think of, let’s say, the first 100’000 tweeted reactions and re-tweets (causing themselves reactions and so on. . .) from the first 10 years after the original tweet stirred up the internet, the markets and the politics . . . what should we do with the 100’001st reaction (maybe by the original twitterer himself?) on the first morning of the 11th year? Should we trust that Twitter, the company, will keep them? In a way like the usenet was “saved” – or rather: *not* saved by Google? So, the first question is: Who should be responsible for the archiving of such ever-changing ‘streams’ of information like a twitter discussion, a personal or institutional Facebook timeline, even a simple discussion in a chat that might be of historical relevance like the protocols of talks at a peace conference? And what about the streams of video conferences over Skype, FaceTime or Firefox hello etc.?

Ok, these last ones we may subsume under the digital document paradigm because we may suppose that these discussions are terminated some day and could be archived then. But there would still be the problem with “streams” in Facebook etc. whose bits and parts are aggregated ad-hoc from very different sources, saved in very different file systems and computers or servers, in some cases surely already distributed all over the world, literally. Where is the ‘digital document’ the archivist should or would try to preserve for the future in these cases? In addition: I, for instance, take part in chat-room discussions regarding software problems that can have been silent for years before the same (or rather a similar) problem or question comes up again and is discussed or solved. Can we be sure that important administrative discussions today do not develop in the same direction in the near future? – Especially: What will happen the longer the more new generations of the personell involved get accustomed to digital tools and use them instead of the old-fashioned e-mails that one could still save as digital files and even print out for archival purposes?

Of course, one may say, these forms of a somewhat ‘fluid’ communication are comparable to talks and ephemeral contacts from the past, and that we should archive only the results (or what we may regard as a stopping point) in some kind of a ‘cutout’ from very long ‘document’ like it was done over centuries. But this, still, maintains the ‘document paradigm’ which is only one view, – and one may add: – the somehow ‘more natural’ view on the digital data at least for humans.

But I guess the ‘digital document paradigm’ (as part of the ‘desktop–folder–file paradigm’) will lose ground to the ‘streaming-like’ forms of communication – and is already doing so on a rapid level, if we look at the media that are used today for communication. I think you may agree that the tweet example from above is not very far from reality, in fact: In this special case from just some weeks ago, everything could have happened exactly the same way I described it . . . if the people in charge with access to the crucial information would not be tied to some sort of administrative discipline – or simply: if they were not used anymore to the traditional ways of information dissemination. At least the last ‘argument’ will disappear with the next generation, the one already preparing for administrative careers.

THE DISAPPEARANCE OF THE PROGRAMME OR ‘APP’

Another problem for archivists arises – from my point of view – from the disappearance of dedicated software programs. Of course, we all still work with our different programs to create office document files, save and edit images or movies, write and send emails, ‘tweet’ or update our facebook timelines . . . But I see a development away from the dedicated software and therefore away from the paradigm: “digital files are created with dedicated software” which we may call the ‘App paradigm’. In fact, some of the programmes on our computers, and surely many of the apps on our handhelds are hardly software programmes in themselves. If they are, this is rather due to the commercial aspects of the software industry than to technical limitations: If we look back to the 1990s there were already office programs like Star Office that suggested to “do everything in one place” containing text, spreadsheet and presentation programs together with e-mail clients and web browser under one uniform user interface. But we can go back even further:

When Douglas Engelbart held his famous demonstration in 1968, which still is a very fascinating documentation, especially when you watch it with the question in mind: Could I do these things on my computer today? – So, when Engelbart gave his ‘mother of all presentations’ he did not ‘open a program’ to create and save a file, and then opened another program to do something else. In fact, he saved text he had written on the screen ‘to a file’ (as he said a few times), but he could jump to any text element (word or link, line of code etc.) ‘inside’ these ‘files’ from any other file he was just working on – without opening any dedicated program to work with one of these files.

Today, if we look at the ‘apps’ on our handhelds we may guess that their development moves into a similar direction: Of course, and only due to the limitations of the software business, we still work with dedicated programs, but in the background these are rather conglomerates uniting different functions like the graphical user interface and its components with other programs or functions providing online connection or the ability to type and display text, to record our voices or images and movies etc. My guess is that in the near future, users will not be willing to accept these procedures: start this program, create a file, save the file, start another program, open another file, copy one element from the first file to the new file etc. Rather, users will expect their computers and handhelds to (as far as possible) automatically recognise the things to do and the data to save. – If we take StarTrek as a very good prediction of IT things to come (the mobile phone as well as the tablet, even the computer ‘understanding’ our voice and answering our questions – ‘invented for StarTrek’ and available today ... if we take this as a prediction, than *Siri* and other assistant programs knowing what to do with our input will be much more common in a few years. — So, where, then, is the program that we could save, e.g., in Vint Cerf’s “digital vellum” to use our ‘digital documents’ in the future or preserve them for future generations?

THE DISAPPEARANCE OF ‘THE’ COMPUTER AND / OR SERVER

We but especially and even more the younger generation(s) is/are already accustomed to their handheld devices and to store their data in ‘the’ cloud. But today, these devices are more powerful than most of the ‘super-computers’ the World Wide Web was built with some two decades ago – and they are surely more powerful than those computers from the beginning of the internet around 1970. So: Why should, for instance, these handhelds communicate at all over dedicated server computers on the internet? As long as I have a stable IP address, my handheld could already be constantly online as a server delivering my documents to the rest of the world over the internet via the built-in mobile version of Apache, for instance.

Already today, I could have all my documents, images, photographs or movies “in the cloud”, i.e. scattered over dozens of virtual servers from dozens of companies – many of these servers surely are ‘virtual’, i.e. distributed over several physical servers, maybe not only across server racks but across server farms or even across several locations all over the world (at least technically this is possible already and could be a usual configuration soon). I still may have all these data on my mobile phone with its 128 GB of memory, but, if it breaks: where are my data then? In my backups at home, ok. And else? I do not belong to the young generation whose life takes place in larger parts via Facebook – where are their data that should be kept for the future? Which database serving as a backend for any sort of ‘files’ somewhere “in the cloud” should be preserved (and how

often) to document my activities, my knowledge and my opinions for the (near) future? Of course, one may say: you are not that important . . . but the Bernd-Kulawikology of the future may see it in a different way. And I, having worked with computers since 1983 may be a rather “good” case for future archivist, because I tend to backup my data on several harddisks in different places. But, you will agree: the average politician producing data worth archiving *will not (want to) care* about all this technical riff-raff . . . and he will trust in his IT staff to keep everything somehow and select the important stuff at the State Archive. But can he trust the “IT croud” and their tools? I agree with Vint Cerf: At the moment – he can’t.

So: where *is* ‘the’ server whose states we or the contemporary or future archivist could try to save regularly with all the files, databases, user interactions, ‘streams’ etc.?

CONCLUSION

To sum up:

From my point of view, not only the ‘digital document paradigm’ will disappear with its imitation of the paper-based document, but also the dedicated programs we use to create and edit these documents and even the dedicated ‘servers’ that run the software and save our documents. What we will have and already have instead in many parts of our digital lives, is a ‘network of floating digital data’ being put together in real-time in something we could compare to a puzzle or mosaic instead of a document – as it was already visible in Doug Engelbart’s presentation 47 years ago. . . . But in the (not so) long run even the ‘digital’ in the sense of Zeros and Ones – will disappear and will be substituted by – e.g. – quantum computers that can handle multiple states.

So, what could or should the archivists of today or in the future do to avoid the ‘digital black hole’? More general: What should we all working in administrative environments do? Or even more general: what should we as private persons do with our data? Of course, every archivist and historian knows (or should know) that private documents from history are of less, the same or even more importance than those coming from governments, companys and other organisations. (And a ‘conspiracy theorist’ might even add: the really important information is not in the public records.)

But, to say it frankly: I don’t see any solution.

Should we trust that digital data – at least those available over the web – are somehow somewhere saved by Google, Facebook and the companies offering cloud services? Of course NOT.

Should we trust that the NSA is saving *every* traffic and digital ‘action’ occuring in the internet? At least: their new storage center in Utah is said to be able to store the entire internet traffic for the next 3-4 years. – But again: of course NOT. Not even if the NSA should be able to save the content of all of our private computers, handhelds etc.

So, then, what could we recommend to the archivist: The only storage medium trustworthy / trustable for the next 100, 200, 500 years that I know of, is paper. Or does anyone who remembers the times of punchcards, punched tapes, magnetic tapes, magnetic harddisks (yes, they are disappearing, too) or laser discs use these today *AND* expects them to be in use in 10, 20 or 50 years – not to speak of 500? Does anyone expect Vint Cerf's "digital vellum" to be stable for more than 20 or 50 years? I don't ...

But the problem is, as I described it above: the disappearance of the digital document will make it hard to save something derived from the digital data streams as a printable document. So, the paper solution is not really a solution – more or less in the same way as the parchment (be it real or in the form of any "digital vellum") or the cuneiform clay tablet would be ... even though the latter one could store information for at least 5'000 years.

So, is there no chance to save at least parts of our information heritage for the future? Do we really live in times that future generations will call the "digital dark age"? (At least, the cynic may say, this would give them the chance to re-discover everything in a new Renaissance ...)

I only see one chance and that is: to focus all available forces from computer science to companies and political administrations *AND archives* (because they have the real experience with long time preservation!) to develop tools, software, computer systems, file formats and everything necessary for long-term storage solutions. While some institutions or persons (like Vint Cerf) start to think about to save at least part of our digital information produced today, in the past (how many databases from the 1980s and 1990s are still usable?) and the near future of the next 10 to 20 years – this is not enough and it comes too late already.

Therefore: Governments, companies, organisations should be involved and made responsible to only use solutions, software, systems, etc. for which a uniform standard of preservation of *any* kind of data has been developed. This development may take up to 20 years, it may cost large amounts of money and manpower – but to lose almost all our digital information or leave its storage to commercial companies, who need to make money with it, or to secret services, who may disappear after a democratic decision ... is no option at all.

Do we have the money, manpower and energy to develop such standards and solutions. Of course we do! Today, they are only dispersed over companies, universities, institutions and used for short-term aims like making money and surveilling the people. I guess, we have to make a decision – and we have to make it soon.