PPTS, PDFS, AND PROPRIETARY FORMATS: ON THE PROBLEMATIC FORMS OF THE DIGITAL

Debbie Ding
Visual Artist and Technologist

The digital divide is often used to describe the chasm between people who have access to technology and those who do not. But is there another digital divide, caused by the problematic forms of the "digital"? Do we have a common understanding of what it really means to "digitise" information?

In this paper, I will examine a number of commonly used digital formats such as PowerPoint (PPT) and Portable Document Format (PDF), and discuss how the process of digitisation is not preservation but the creation of a new digital object.

A brief history of the term "digital divide"

The earliest use of the term "digital divide" can be traced back to the mid-1990s when it was used by the United States administration and journalists to describe the educational, economic, and social inequalities between those who have access to a computer and the internet and those who do not.

In July 1995, the newly formed National Telecommunications and Information Administration (NTIA) in the United States published a series of reports and NTIA

Administrator, Larry Irving defined the "digital divide" as "the divide between those with access to new technologies and those without" (National Telecommunications and Information Administration 1999). The White House also began to use the term. Speaking about a mobile "cyber-education" computer lab in a truck, then-Vice President Al Gore described it as "rolling into communities, connecting schools in the poorest neighbourhoods, and paving over the digital divide" (quoted in Williams 2001, 2).

The core idea behind the use of the term "digital divide" was the growing awareness that access to technology cannot be assumed to be automatic or universally applicable, and so governments must work to improve infrastructure and ensure the even penetration of information communication technologies.

In Singapore, much has been done to address the perceived "digital divide". Starting with the first Masterplan for IT (Information Technology) in Education in 1997, several IT masterplans have been implemented as a comprehensive strategy for ensuring "every child (would be) proficient in the use of computers and benefit from learning in an IT-

enriched environment" (Ministry of Education 1997). Singapore has built up excellent hard infrastructure over the years. The IMDA (Infocomm Media Development Authority) has also developed sophisticated and targeted programmes to reach out to segments of the population that are less internet-savvy. For the last few years, Singapore has been leading the World Economic Forum's Global Information Technology Report in its list of most tech-ready countries (World Economic Forum 2015).

The rise of the term "digital divide" marked the progression from the utopic 1980s "cyberbole" of the "global village" to concerns about the circumstances which limit people's access to technology and data.

As the use of technology becomes more and more pervasive, there is need for further progression in our understanding towards a critical awareness that when something is digital or "digitised", there also may arise a "digital divide" of a technical kind – such as the problem of whether the available digital data or digitised material is machine-readable, structured, in an open or proprietary format, and whether the conditions of its release are complicated by other interests.

Defining the "digital"

What does it really mean when something has been "digitised", or is digital? According to the Oxford English Dictionary (OED), "digital" refers to "using, or storing data or information in the form of digital signals". On the OED's "Dictionary Facts" webpage, it also playfully estimates that it would take an equivalent of 120 years to "key in" all twenty volumes and 21,730 pages of the 1989 Second Edition of its dictionary in order to convert it to a machine-readable form, and it would take another 60 years for a person to proof-read all the entries (Oxford University Press 2016).

As the term "digital" does not indicate whether the "user" is man or machine, there is potential for confusion as being "digitally accessible" is not necessarily synonymous with being "machine-readable". For example, a digitally scanned document with text and graphics within it may become "digitally accessible" in the sense that we might be able to view a scanned digital image of it on a computer, but the computer will not be able to "understand" the information within unless the text itself has been machine-encoded and structured.

Commonly used digital formats such as PDF, Joint Photographic Experts Group (JPEG), Document (DOC), and Hypertext Markup Language (HTML) are not machine-readable, meaning that while the computer is able to visually display and store the digital data, it is unable to read or understand what the data represents. This may directly affect or limit the uses that different parties can derive from the information, and it would also require more work for this data to be extracted and processed by a machine.

The Problem with Powerpoint: Human misreadings

The problem with the digital is more than just issues with machine-readability. Even digitally accessible formats that are supposed to be human-readable are not without their problems, as it does not necessarily mean that all humans can write them, read them, or comprehend them even if they are able to access them digitally.

PowerPoint, the highly ubiquitous slideshow presentation software developed by Microsoft, is one example of a human readable digital format that is used by people around the world to produce visual aids that help them communicate ideas or conduct briefings and presentations. The conference at which this paper was

presented, like almost all conferences in the world, also uses PowerPoint.

In an oft-quoted example in information design studies, American statistician Edward Tufte famously argued against the use of Microsoft PowerPoint in corporate and government bureaucracies because PowerPoint's "low resolution" cognitive style of presenting critical technical information over several hyper-rationalist levels of "bureaucratic" hierarchy may have adversely affected the quality of NASA's engineering analysis – contributing to the 2003 crash of the space shuttle Columbia:

> During the January 2003 spaceflight of shuttle Columbia, 82 seconds after liftoff, a 1.67 pound (760 grams) piece of foam insulation broke off from the liquid fuel tank, hit the left wing, and broke through the wing's thermal protection. After orbiting the Earth for 2 weeks with an undetected hole in its wing, the Columbia burned up during re-entry because the compromised thermal protection was unable to withstand the intense temperatures that occur upon atmosphere re-entry. (Tufte 2006, 162)

Tufte argues that the format of the PowerPoint slide resulted in engineering reports that visually downplayed the potential risks caused by the damage to the left wing. Without any prior guidelines on how to communicate dense and serious technical analysis involving the survival of the shuttle into the laconic format of a "powerpoint pitch", Boeing Corporation engineers ended up summarising their detailed studies into acronyms, clipped phrases, vague pronoun-antecedents, and bullet points with unnecessary hierarchies. Large fonts resulted in typographical orphans, leaving crucial numbers and their occasionally inconsistent units on different lines.

In another study published in the journal Technical Communication involving a survey of 1,014 participants including students and employees from different fields, it was found that 36 per cent of the preparation time for the average proposal was consumed by design and animation work by people without formal graphics training (Thielsch 2012), meaning that the time spent on producing the digital presentation file was taking up significant time that could have been used to better develop and focus the content of the proposal.

The invasion of pitch culture into so many other areas of work, even the most technical analysis and the use of the digital presentation slide format therefore presented itself as a divider, or obstruction to the understanding of NASA's upper level executives.

The Problem with PDFs and Proprietary Formats: Machine misreadings

When the question shifts to tasking machines with reading our digital formats, what are the implications when human-readable digital formats cannot be read by machines?

Another commonly used digital file format worth taking a closer look at is the PDF or Portable Document Format. First launched in July 1993 and made free to use later that year, it was one of the first digital document file formats that could be shared electronically while promising to retain all the elements of its original formatting across platforms.

The technology for PDF came about from a side project operated from within Adobe by Adobe co-founder John Warnock who wrote:

> Imagine being able to send full text and graphics documents (newspapers, magazine articles, technical manuals etc.) over electronic mail distribution networks. These documents could be viewed on any machine and any selected

document could be printed locally. This capability would truly change the way information is managed. Large centrally maintained databases of documents could be accessed remotely and selectively printed remotely. (Warnock 1991)

Today, most conference papers, reports, and important documents are circulated in the PDF format, many of which would have been produced at a time before digital preservation was recognised as a concern for the future.

PDF itself began as a proprietary format that was controlled by Adobe, but it was released for free and officially released as an open standard on 1 July 2008. However, there are a few proprietary technologies defined only by Adobe, which are not supported by third-party implementations of PDF readers and editors, so when PDFs are published using these non-standardised features, these may present some issues with accessibility.

A list of potential preservation risks that could occur with the use of PDF can be found in Inventory of Long-term Preservation Risks by the National Library of the Netherlands. For example, it notes that certain features in PDF files such as embedded data that may rely on external applications which may not be available in the future, problems with fonts (which may be required for the display of different languages), and the use of lossy image encodings may compromise image quality of these records in the long run (van der Knijff 2009).

In 2003, Guardian's computer editor Jack Schofield formulated "Schofield's First Law of Computing", which says to "never put data into a program unless you can see exactly how to get it out." While not suggesting that all proprietary formats are inadvisable for use, it is a reminder

that having a wealth of digitised material is no use if your data is trapped in software formats that you might have a problem opening in the future.

Digitisation is not preservation but the creation of a new digital object

Many approaches towards data today seem centered around man's ever-increasing technological ability to capture, consolidate, compress, and transmit a large volume or velocity of data. Amongst all this data, there is often a distinction made between digital objects that are "born digital" as opposed to those which have been "digitised".

To some extent, it is true that there is a distinction to be made between an object that began its life as a digital object, and a digitised object, which came into existence in order to record some of the characteristics of a physical object.

However, this distinction made between born digital and digitised material may not be helpful when one takes into consideration that unlike the digitised object, the born digital object in our cultural spaces may not always be created with a planned use or planned mode of preservation in mind, whereas we always remain concerned about the fidelity of the visual appearance of the digitised object to its original object. It cannot be taken as a given that the digital object is already in a format that is appropriate for long-term digital preservation.

It must be made clear that digitisation itself is not an act of preservation but the creation of a new digital object. Like another kind of digital divide, since the process of digitisation and preserving digital objects has the potential to radically alter our texts, images, and cultural capital; this underscores the importance of continually interrogating the way in which we digitise, display, and make documentations in the fields of arts and culture.

References

1.   Knijff, Johan van der. 2009. *Adobe Portable Document Format: Inventory of Long-term Preservation Risks*. Koninklijke Bibliotheek, National Library of the Netherlands. http://openpreservation.org/system/files/PDFInventoryPreservationRisks_0_2_0.pdf.
2.   Ministry of Education. 1997. *Speech By Radm Teo Chee Hean, Minister For Education and 2nd Minister for Defence at The Launch of the Masterplan For It in Education on Monday 28 April 97 at Suntec City at 10 am Opening New Frontiers in Education with Information Technology*. https://www.moe.gov.sg/media/speeches/1997/280497.htm.
3.   National Telecommunications and Information Administration, United States Department of Commerce. 1999. *Falling Through the Net: Defining the Digital Divide*. 8 July. https://www.ntia.doc.gov/legacy/ntiahome/fttn99/contents.html.
4.   Oxford University Press. 2016. "Dictionary Facts." *OED Oxford English Dictionary*. http://public.oed.com/history-of-the-oed/dictionary-facts/.
5.   Thielsch, Meinald. 2012. "Use and Evaluation of Presentation Software", *Technical Communication* 59 (2): 112-123.
6.   Tufte, Edward. 2006. *Beautiful Evidence*. Cheshire, Conn.: Graphics Press.
7.   Warnock, John. 1991. "The Camelot Project." *Adobe Blogs*. http://blogs.adobe.com/acrobat/files/2013/09/Camelot.pdf.
8.   Williams, Kate. 2001. "What is the digital divide?" In *Proceedings of the Digital Divide Doctoral Students Workshop*, 1 August 2001. Ann Argor, MI: University of Michigan. http://people.ischool.illinois.edu/~katewill/williamsd3workshop.pdf.
9.   World Economic Forum. 2015. *Global Information Technology Report* 2015. http://reports.weforum.org/global-information-technology-report-2015/.