An Exploration in Online Storage on Paper

Berend Nordeman

Media Technology, LIACS, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
berend@nordeman.nl

ABSTRACT

As the Internet is becoming of increasing importance in our daily lives, the amount of personal data stored online also increases. The way data is stored is in stark contrast with having information on paper. This research explores the combination of online storage with physical paper. Previous research on the combination of paper with computers mainly focused on data entry and human-computer interaction. In this research a webserver using a paper file system is described and an implementation of that system is made. After building and using the system we conclude that a paper server is not of any daily practical use but does result in new insights.

1. INTRODUCTION AND MOTIVATION

In 2011 a company specialized in online data storage introduced a new and even more secure backup medium: paper [1]. Instead of humming servers the data center was filled with large quantities of paper and barcode printers. It turned out to be an April's Fools joke. However, this joke has a serious side.

Our current daily life is filled with technology. It is not only helping us with difficult tasks, our society is depending on it. The most recent trend is that businesses and individuals are always connected to the Internet using fixed and mobile devices. The Internet is not only used as a medium to communicate between devices, it is also becoming the center of our digital lives. We communicate via online services like email, chat and social networking sites and use online storage services for our documents, pictures, videos and other files. They are often referred to as 'the cloud'. Online storage has a lot of advantages. For example: the files are almost always available, at home and on the go, as long there is a connection to the Internet. Also, there is a virtually unlimited amount of storage and the files can easily be shared with other people. There are also some disadvantages however. Because of the virtual nature of online storage it isn't always clear where the files exactly are. They could be everywhere due to backups, redundancy and replication. Access to the files is managed by an external party and there is usually no or little information about what is happening with the files. Even deleting a file doesn't guarantee it will be gone forever. This all affects the amount of control a user has on its files.

This is in stark contrast with having information offline, on paper. There are many differences between online storage and information on paper. Compared to virtual media, this physical medium is slow, expensive and can not be manipulated like virtual media [2]. It is also transient, not

	digital	analog
offline	USB-drives	paper
online	The cloud	?

Table 1: Examples of media categorized by the properties offline, online, digital and analog.

interactive and bound to its physical location. It is intuitive, portable and easy to use since people are used to work with paper. Although it might be a disadvantage that it is bound to a physical location, it could also be an advantage because *it is where it is*.

Floppy disks, USB-drives, portable HDDs and memory cards are a combination of these two. Because of the digital nature of these devices, files can easily be manipulated and copied like files stored online. However, they are still bound to the physical location of that device and losing the device means losing the file.

Each of the three categories described above are an analog or digital medium and are either online or offline. When put in a table, a fourth category is introduced: analog and online (Table 1).

This paper explores what could possibly be placed in this fourth quadrant. Since there is no direct need for a solution to a problem, an artistic approach is used, even though this could result in more questions than answers. By enlarging properties of topics covered by this research to extreme sizes, new insights can be gathered. Because half of the research is about the analog, physical domain, a physical installation is considered appropriate.

To start the exploration we first take a look at research done in a similar direction. In the third section we describe a theoretical system of which an implementation is made (section 4). The results of the implementation are discussed (section 6) followed by and a conclusion (section 7).

2. RELATED WORK

There has been a lot of research on using paper as an input or output medium in combination with computers. When paper was the most important carrier of information for businesses there was a large industry for filing systems. After most companies switched to computers, converting all these paper documents to their digital equivalents was a major task and different systems for converting were developed. These digital filing systems usually consist out of a scanner to digitize the document, optical character recognition to convert the image to digital text, and a database to store the collected information. Rao *et al.* describe a system called *Protofoil* [3] which essentially functions as an electronic document filing system using a scanner or fax as input. Because of this documents can be added to the database from all over the world as long there is a telephone connection. A special form that acts as a paper interface is added to the document to be scanned. In this form check boxes and text fields can be filled with information about the document or operations the computer has to perform on the document. The concept of a form to control a computer is called a paper user interface. This concept, implemented in a system called XAX, is written about by Johnson *et al* [2].

Some of the research focuses on using paper as a tangible user interface. Examples of this are *iCandy* [4] and *Palette* [5]. In both projects small paper cards are used as a representation of some virtual data. The cards contain links in the form of a barcode to an music album and presentation slides, respectively. These projects show that physical objects are intuitive to users and can enrich the way people use computers. However, they are one-way and only offer interaction with a computer. *Quickies* [6], *The Designers' Outpost* [7] and *PapierCraft* [8] allow free form user input that is digitized and used by software. In the case of *Quickies* information written on notes is interpreted and processed. *The Designers' Outpost* uses physical sticky notes as input. These notes are digitized and can be virtually linked together using gestures on a physical smart board.

PapierCraft is an example of combining the advantages of physical paper with the easy editing of virtual documents. This is implemented by merging annotations on paper with a digital copy and using drawn commands to edit the copy. Heiner *et al.* describe a combination of a paper notebook and a digital PDA called *Paper PDA* [9]. This research is based on the observation that a paper notebook can be synchronized with a digital one by scanning the pages, recognizing the contents, combining that with digital information and printing it as a new page. This close interweaving of physical and virtual shows the advantages of both virtual and physical media and how the best of both can be combined.

A project that tries to remove every user interface is the *Tableau* [10]. It is a nightstand with a drawer that makes digital pictures physical. The pictures are retrieved from a Twitter feed and printed on small papers. It is also possible to put a note in the drawer which will be photographed, uploaded and posted on a Twitter feed. This device automates most actions needed to communicate over the Internet. It can store physical pictures, but retrieving images is limited to the last printed image.

Another project that uses Twitter is *Murmur Study* by Christopher Baker [11]. It consists out of 30 thermal printers that monitor status updates on Twitter and Facebook and print those containing common emotional utterances. Baker makes these texts physical to show the amount of personal expressions that are visible for the whole world and are there to stay. This to make viewers of the artwork aware of the virtual trails they leave behind.

Email has always been printed a lot. Although it was an April Fools' joke, Google launched *GMail Paper* [12] as a service to print all of your emails and send them to you via the postal service. The most notable element of this joke was that it appeals on the need to have emails on paper.

Albeit no actual machine but only a theoretical description, the Memex [13] described by Vannevar Bush deals with storage of information on physical information carriers and still being able to share it with other systems.

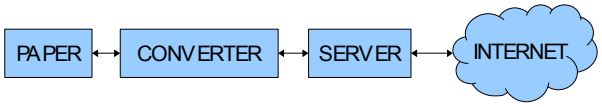

Figure 1: Schematic view of theoretical system

Most projects that use paper as input for computers are aimed at human-computer interaction or as a method to enter data into the system. The projects that use paper also as an output medium focus on 'interactive' paper. None of these projects use paper just for storing data as one would on a CD-ROM or USB-drive. In the next section we describe a theoretical system that features this property.

3. THEORETICAL SYSTEM

To fill the vacant space in Table 1 we propose a combination of paper with 'the cloud': a system that makes data stored on paper available via the Internet.

The system should use a type of paper as storage medium which users are familiar with to increase the usability. Since the storage medium is physical and the input and output are virtual the system should contain a converter. Depending on the form of the paper a feeding system should be added. There should also be a method to select the right paper and data. The paper should be easy to handle by the user and the data should be applied to the paper in a user readable format. These separate systems should all be connected to a central unit that communicates with the Internet, acting as a server.

The software on the server has the task to make the data on the paper available. Since virtual duplicates of the original should be avoided, a common web server where users can unlimited get documents from is no option. Instead, a method should be used to transfer data to a receiver while being able to remove that data at any time and being able to prohibit copying and spreading of that data.

When saving a file to the system the original virtual file and all of it's instances should be

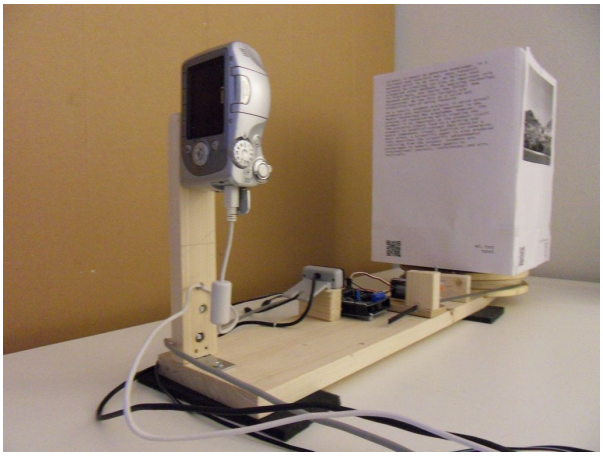

Figure 2: The 'paper' and 'converter' parts as shown in figure 1

erased.
Deleting files from this system is done by removing the paper containing the file from the system.

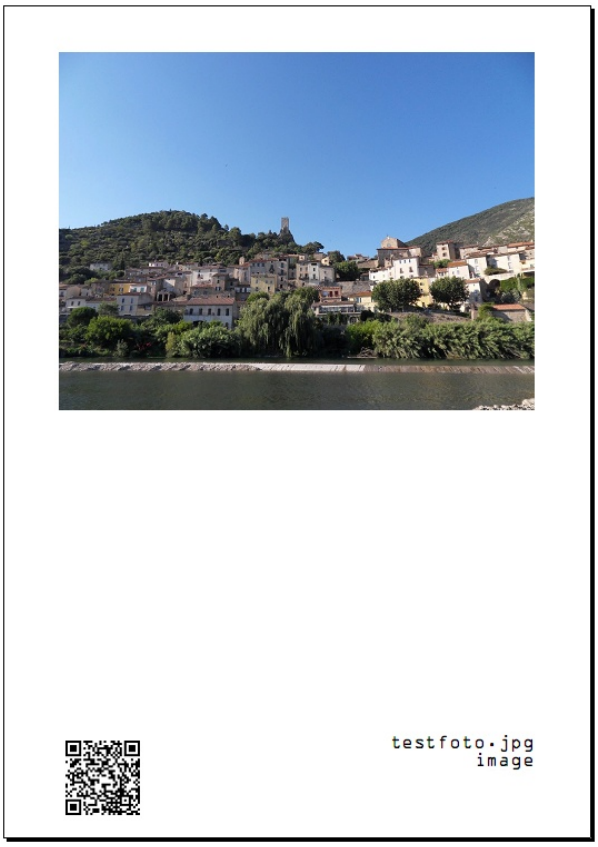

Figure 3: Example of a file used by the system. The sheet contains the file 'testfoto.jpg'.

4. IMPLEMENTATION

This section describes an actual implementation (fig. 2) of the theoretical system as described in the previous section. There are many ways the theoretical system can be implemented. One could focus on building a system that is usable in an office environment or focus on consumers at home. There could also be requirements for speed, capacity and dimensions and many other factors. This implementation focuses on the possibility and consequences of combining paper with online storage and should be considered art rather than a system for daily use.

The system uses sheets of A4 paper as the storage medium (fig. 3). This format has been chosen because of its ready availability and the ease to print onto. The bottom part of each sheet is reserved for metadata in human readable form and encoded as an QR-code. The metadata contains the filename and filetype. Of each sheet only a single side is used so that files are also physically separated and to make the mechanical handling of the sheets easier. Sheets are created manually using a word processor. The QR-code is generated using qrencode-win32 [14] at error level H and contains the filename and filetype separated by a colon. The size of files is limited by the size of the data part of the sheet and the detail of the printing. In this implementation a text file is limited to around 300 words and an image is limited to around 3000 pixel in width and height. This could be increased by using a higher resolution camera or other apparatus to convert from analog to digital. The capacity of a text file could also be increased by using higher quality optical character recognition software.

The system has a storage capacity of four sheets so it can store up to four files. The sheets are placed on the sides of a square carousel and these are the 'paper' part of the theoretical system. The carousel can be rotated with a motor in steps of 90º so that each of the four files can face the cameras. A single step takes around five seconds.
On one side a small webcam is directed at the carousel and focused on the position of the QR-code containing the metadata. A Canon PowerShot SX100IS digital camera is directed at the part of

the sheet containing the contents of the file. The camera is the 'converter' part of the theoretical system and both are connected to the server.

The server consists out of a combination of software running on a common computer and Microsoft Windows XP as operating system. The Apache HTTP daemon with the PHP module is responsible for communication with the user. There are no files stored at the webserver so every request will result in a 404 Not Found error message. A modified 404 error page written in PHP is used to broadcast the requested filename and a random string via the OSC protocol. The page will then check for a file with the random string as filename for a minute every five seconds. If the random named file has not been found after a minute it will return a 404 Not Found message. During this minute the system will scan the QR-codes on the sheets for the requested filename. This process is controlled by a program written using the Processing Development Environment and starts with the incoming OSC message containing the filename and random string. The Processing sketch reads a video frame from the webcam and decodes the QR-code of the sheet currently facing the cameras. If the filename in the metadata does not equal the requested filename the sketch sends a 'next' signal to an Arduino Diecimila microcontroller with an Adafruit Motor Shield [15] on top. This shield controls the motor and servo that rotate the carousel. The carousel is rotated 90º so that the next sheet faces the camera. This also happens when the side of the carousel is empty and no QR-code could be found. The process repeats for each side of the carousel or until the requested file is found. If the requested file is not found by the sketch the PHP script will time out and return a 404 Not Found error message. If the requested file is found the sketch instructs the BreezeSys PSRemote software [16] to take a picture of the data part of the sheet using the digital camera. If that file is a text file, the image is converted to a digital text file using the optical character recognition program GOCR [17]. The file is stored on the webserver using the random string as the filename where the PHP script will find it and return it as a in answer of the request. After transmitting the file is deleted from the webserver. The time between an incoming request and returning the file depends on the filetype and the location on the carousel. The times range from 7 seconds if the requested file is facing the cameras to 50 seconds for large text files on the last position on the carousel.
The system is approximately 37 cm height, has a width of 21 cm and a length of 60 cm. It is constructed mainly out of wood and foamboard.

5. DISCUSSION

The system introduces some new situations and new possibilities which will be discussed in this section. First of all, it is not a perfect implementation of the theoretical system. This implementation focuses on the use of paper as a storage medium and does not intend to be a system for everyday use. It has been used to test and demonstrate the principle of using paper as an online data storage medium.
At some points the implementation differs from the theoretical machine. The core elements of the theoretical system (paper, converter, server) have been implemented successfully. However, the protection of virtual files send to the user can't be guaranteed using current technology and has therefore been omitted. This resulted in the use of a common webserver with a modified 404 Not Found error page.
Since the storage capacity of the system is limited there was no need for a large amount of data sheets during the research and therefore the data sheets were created manually. This process could easily be automated if required. There are more improvements that could be done to the system. Pre-printed QR-codes on a sticker containing a known filename could turn each sheet of paper into a data sheet. Using a higher quality OCR program it would even be possible to replace the QR-code with a handwritten filename and metadata. In the case of a text file it would even

be possible to open a file by entering some words that occur in the document stored in the carousel. This would remove the need for creating a data sheet.

Using the same software the system could have been extended to support other filetypes, including binary files. This functionality has not been added because it was not needed in demonstrating the concept. The content of the filetypes supported in this system, text and pictures, are user readable. The content of a binary file is hard to understand because of the encoding necessary. Even though a file is not readable by a user, it still has the advantage of being handled as a physical object.

This system does not aim to be a combination of the best of both worlds as it also includes some of the disadvantages of both.

An earlier prototype of the machine had a capacity of 100 sheets of A4 paper. It consisted out of two stacks of paper with a suction pump on an arm to move papers between the stacks. Because the automated handling of the paper did not function as expected this design was abandoned. We also came to realize that the focus was too much on creating a machine for daily use instead of a machine that would serve the research. This realization resulted in a simplified design and used a carousel with four sheets for data storage. The carousel is the least space-efficient arrangement but it shows that paper uses a lot of space compared to the amount of data that can be stored on it, emphasizing the limitations of the medium.

In its current form, the system is clearly not suitable for large files because they would have to be spread over multiple sheets. This would impair the physical handling of the files.

Partly due to the bulky mechanical nature of the system it is very noticeable if a file is requested from the server. This raises the awareness of a file being accessed but could also be experienced as annoying.

Because for most people the availability of free online storage is larger than their demand there is no need to be selective about what to store. The storage on paper is much more expensive. This imposes a more critical look on whether or not a file should be made available at all.

From an artistic perspective the system addresses the question “where's my data?” It shows that having control over it also requires more involvement with the data and its distribution. It requires to strike a balance between having full control over it and the convenience of outsourcing.

The system also shows that there is a small gray area between online and offline. Files stored on the server – the sheets on the carousel – are available through the Internet, so they are online. However, they need to be photographed first. In that sense the Eiffel Tower is online too, because everyday lots of pictures of it are being uploaded to online storage services. This is a topic that could be researched more.

6. CONCLUSIONS

The goal of this project was to research the combination of analog with online by combining paper with online storage. The research has resulted in the description of a theoretical machine and an actual implementation of a machine that makes information stored on paper available through the Internet. We have found no direct practical use for such a machine. However, it brought us many new insights. The system shows that files can be online and offline at the same time, how files can be kept offline until they are requested and shows some advantages and disadvantages of using paper as an online storage medium.

7. ACKNOWLEDGMENTS

The author wishes to thank Maarten Lamers for his help in the initiation of the project, Hanna Schraffenberger and Bas Haring for commenting on the preliminary paper and Peter van der Putten for supervising this project.

8. REFERENCES

[1] Livedrive Internet Ltd. (2011, April 1). Livedrive announces industry leading, game changing technology for online backup and data storage.

http://blog.livedrive.com/2011/04/livedrive-announces-industry-leading-game-changing-technology-for-online-backup-and-data-storage/

[2] Johnson, W., Jellinek, H., Klotz, L., Rao, R., and Card, S., Bridging the paper and electronic worlds: The paper user interface, *Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems*. IOS Press, Amsterdam, The Netherlands, 507-512.

[3] Rao, R., Card, S. K., Johnson, W., Klotz, L., and Trigg, R. H., Protofoil: Storing and Finding the Information Worker's Paper Documents in an Electronic File Cabinet, *Human Factors in Computing Systems* (CHI'94), pp. 180-186, Boston, MA, Apr. 1994.

[4] Graham, J., and Hull, J. J., iCandy: a tangible user interface for iTunes, *CHI '08 extended abstracts on Human factors in computing systems* (CHI EA '08). ACM, New York, NY, USA, 2343-2348.

[5] Nelson, L., Ichimura, S., Pedersen, E.R., and Adams, L., Palette: A Paper Interface for Giving Presentations, *Proceeding of the CHI 99 Conference on Human Factors in Computing Systems* (CHI '99), ACM Press, pp. 354-361, May 1999.

[6] Pranav, M., and Maes, P., Augmenting Sticky Notes as an I/O Interface. *Universal Access in Human-Computer Interaction. Part 2: Intelligent and Ubiquitous Interaction Environments*. Springer Berlin/Heidelberg, 2009. 547-556-556.

[7] Klemmer, S. R., Newman, M. W., Farrell, R., Bilezikjian, M., and Landay, J. A., 2001, The designers' outpost: a tangible interface for collaborative web site design, *Proceedings of the 14th annual ACM symposium on User interface software and technology (UIST '01)*. ACM, New York, NY, USA, 1-10.

[8] Liao, C., Guimbretière, F., and Hinckley, K., PapierCraft: a command system for interactive paper. *Proceedings of the 18th annual ACM symposium on User interface software and technology* (UIST '05). ACM, New York, NY, USA, 241-244.

[9] Heiner, J. M., Hudson, S. E., and Tanaka, K., Linking and messaging from real paper in the Paper PDA, *Proceedings of the 12th annual ACM symposium on User interface software and technology* (UIST '99). ACM, New York, NY, USA, 179-186.

[10] Kestner, J. A., 2010. *Social networks for lonely objects.* Masters Thesis, Massachusetts Institute of Technology. 89 p.

[11] Baker, C. (2009) Murmur Study *http://christopherbaker.net/projects/murmur-study/*

[12] Google Inc. (2007, March 30), Introducing Gmail Paper *http://mail.google.com/mail/help/paper/*

[13] Bush, V., As We May Think, *The Atlantic Monthly*, 176(1):101-108, juli 1945.

[14] qrencode-win32 (2011, September 14) *http://code.google.com/p/qrencode-win32/*

[15] Motor Shield - Arduino motor/stepper/servo control (2011, August 8) *http://www.ladyada.net/make/mshield/*

[16] PSRemote - Canon PowerShot Remote Capture Camera Software (2011, September 6) *http://www.breezesys.com/PSRemote/*

[17] GOCR (2011, July 21) *http://jocr.sourceforge.net/*