Information: an obstacle for human rights?

Serguei Chabanov
Serguei Chabanov and Alturing Bt. provide development and maintenance services for EUMAP.ORG website since its creation.

Introduction

The human rights establishment generally perceives the Internet in a positive light. It breaks barriers, we hear, it makes publication and reproduction of information easy, it is an excellent communication tool. Websites, blogs, forums and chat-rooms devoted to the issue are prolific.

Without denying an undoubtedly positive role of the Internet, this paper highlights some of the problems that distinguish it from other media. A number of difficulties with preserving, promoting and verifying information on the Internet can severely undermine its impact and should discourage the human rights community from completely abandoning traditional means of distributing information and communication.

Digital publications: short life, high maintenance

One common misconception about digitalised information is that the digital form somehow makes the preservation of content trivial. Unfortunately, this is not the case. Digital information cannot exist outside of physical media, and the physical media available for this purpose are spectacularly short-lived and volatile when compared to a printed book. There are usable prints of Guttenberg's Bible (15th Century), and even an edition of The Diamond Sutra, printed in 868 in China using wood and clay, is still around. In comparison, an average consumer hard-drive is guaranteed to work for one or two years, and if a CD-Rom was not made using gold for a reflection layer, it will not last for the promised 100 years, either, but closer to 10-20 due to the degradation of the dye and aluminum used. Optical media are sensitive to light and humidity, magnetic media are sensitive to heat and magnetic fields, hard-drives are also sensitive to shock and vibration. To put it simply -- digital media tend to be rather frail in their physical incarnations.

The IT industry itself has little interest in long-term data storage. Faced with a never-ending need to deal with more and more information, it is concerned instead with expanding storage capacity. In other words, a hard-drive will likely become too small to be useful even before it gets a chance to break. Indeed, it often happens that the digital storage medium, volatile as it is, functions longer when it is supported by other hardware and software. For example, it is possible that 5.25' floppy disk from ten years ago may still contain valid information now, but how can it be read, if hardware that previously supported such disks is not longer there?

The speed, with which information to be processed accumulates, also makes it desirable to regularly change the format in which data are stored, and not only the storage medium. Together with the increasing sophistication of software, various formats of data storage evolve at a steady fast pace. Eventually, some formats become unsupported by commonly available software, and to remain accessible, data need to be converted to a new format, which in its turn can turn into a non-trivial task if automatic procedures do not work.

In this situation, the reliability of the storage medium itself is as important as the effort involved in duly converting data through continuously evolving media and formats. In case of electronic distribution, this applies even to information which has already been delivered to its recipients, unless they bother to print it, and even then it is not the same as merely placing a hard copy at the public library. Thus, unless it remains publicly available more or less continuously, digital information cannot be considered publicly available at all, and once it stops being maintained for the public, any traces of it will quickly disappear. (Caution for human rights activists who have not revisited, for instance juridical files, for several years, waiting for court to hear their appeal â€“ the evidence may no longer be at their disposal!)

With large collections of information, such as a documentation database or even a personal mp3 collection, there is an additional task of providing efficient means of accessing it. Any collection needs to be organised in some way for navigation, and indexed for searching, updated and otherwise maintained. Addition of new entries requires manual work, and already existing entries may need to be re-classified because of ever-growing amount of data. Any software that assists in these tasks places further requirements on data formats, and evolves through its own development cycle, with updates, upgrades, etc.

Finally, a public collection of information, such as an Internet database, needs to be kept online, which involves mounting traffic fees and security doldrums from viruses, worms, "script kiddies" [1] and other Internet biota.

All this makes electronic publication a rather different affair compared to a paper print. Instead of one good mass-mailing event, a continuous effort is required to maintain public information. As various projects and initiatives come and go, so do their websites and hosted publications, eventually perishing without an (virtual) trace. There are at present no public Internet libraries one could submit their publication to and reasonable expect it will stay there for posterity.

The above issues are addressed not to scare users from entrusting data to the Internet. Rather, it is to warn against over-reliance on volatile media when preserving vital information. It is possible to spool, share, and pre-empt maintained publications between various projects and initiatives, even develop common libraries sustainable over time, but it requires a serious effort.

Missing identity

Besides troubles with storage and maintenance, digital information is vulnerable in a different, perhaps more serious way. In sharp contrast to other media, there are practically no entry barriers to publishing on the Internet (or CD-Rom, for that matter), and very little in the way of ensuring authenticity of the contents or the source of the information.

Ensuring authenticity of a document involves verifying two things: that the document indeed came from where it claims to have come, and that its content has not been changed since. Solutions that make tasks trivial exist in the form public key cryptography [1] but in practice remain largely ignored by users. One could ask, but cannot one always go to the original website and download the document again? Perhaps, if one knows an original website, and if the original website has also kept the document the same and readily available. However, online publications keep changing, and our knowledge of trusted websites is limited to a few we use regularly. The public, people who are likely to find out about your website through google or wild guessing, have very little reason to trust it. Which of the following websites, for example, are what they seem to be: http://www.whitehouse.org/, http://www.whitehouse.net/, http://www.whitehouse.com/ or http://www.whitehouse.gov/ ? Which website to trust: http://www.aljazeera.com/ or http://english.aljazeera.net/ ?

In the absence of reliable mechanisms of authenticating the information, the ease of its reproduction is as much an advantage as a vulnerability. When published information is intended to have any sort of public impact, it can be effectively circumvented by strategically arranging misleading information. In fact, one of the ways to disrupt the distribution of some file over peer-to-peer network (file-sharing) is quite simply to distribute more and more dysfunctional copies, hoping that the one copy you're interested in would get lost among them and other users will eventually give up on the idea of sharing it.

In conclusion

As long as little attention is given to the problem of continuous availability and authenticity of information and its sources, the Internet is far from being a 100 percent reliable tool for communicating information to the public or gathering any data about it. The situation does not need to remain such indeed, the Internet can be trustworthy enough, for example, for online banking. Some of many possible ways to improve it could be addressed as follows:

- Promoting a certain culture and good practices in dealing with digital information and media, and raising awareness of its problems;
- Developing trust in a project and its electronic publications outside of the Internet - printed ads, presentations, etc;
- Developing partner networks which could provide common level of trust to all users;
- Creating a pool of digital resources maintained by various projects together could be a way to cut maintenance costs and ease preemption.
- Periodically publishing most worthy resources on paper to ensure they remain after the website is gone.

Footnotes

[1] A slang term for amateur hackers who break computer systems just for fun or for hooliganism.

[2] See http://en.wikipedia.org/wiki/Public_key_cryptography/