

Universal Access to Information

By HAL R. VARIAN

In 2003, the world produced about 800MB of information for each man, woman, and child on earth [2]. Much of this information, such as supermarket scanner data and the like, is pretty dull. But some of it, such as the material contained in books, magazines, newspapers, movies, music, and family photos, is potentially of great interest to people.

Peter Lyman and I estimate that well over 90% of information currently produced is created in a digital format, and anticipate this percentage will increase substantially in the future. At the same time, much existing content currently available only in physical formats will soon be digitized. The Internet Archive has already digitized many

gigabytes of audio, video, and textual material. Google recently announced it intends to digitize 15 million library books in the next few years.

These trends suggest that the most useful information will be available in digital form within a decade. The entire corpus of published printed material produced in a year, including books, newspapers, and periodicals occupies between 50TB–200TB, depending on the compression technology. This amount could now easily fit in a refrigerator-size disk array.

So it is now becoming technologically

feasible for all recordable information to be accessible by everyone on the planet. The barriers to accessibility are not technical in nature; they are social, legal, and economic.

First, there is the problem of literacy. There are approximately 900 million illiterate adults in the world, two-thirds of them women [4]. Information, or at least written information, cannot be successfully delivered to those who cannot read. More generally, the usefulness of many sorts of information depends on education of the recipient, and we have a long way to go in this area.

Universal access to all the world's information is technologically possible now; the missing piece is the legal infrastructure that will provide the incentives to make such access economically viable.

Second, there is the problem of infrastructure. Here, Moore's Law is working in our favor, and it appears that it will be possible to build very inexpensive information-access devices. Even today one can buy a TV+DVD player for well under \$100. DVDs hold approximately 4.5GB of information, which is equal to the textual content of 4,500 books. It is possible to put all the textbooks used in an entire university on a few DVDs. Inexpensive technologies such as variations on personal digital assistants could be used to access such material.

By far the biggest problems are the legal and economic issues. How will the costs of universal information access be covered?

We have a number of business models to price and pay for infrastructure. The tricky part is how to pay for content since many creators of works want to be paid for their creations. Copyright has been a workable, albeit imperfect, solution to the problem of compensating authors and other creators for their works.

The problem with copyright is that it uses a tremendously inefficient system for discovering and acquiring rights. Unlike patents, copyrights do not have to be registered to be enforced. Virtually everything you create is automatically copyrighted. This means that anyone who wants to reproduce your works must seek you out and request permission to do so.

This is a terribly cumbersome procedure. It would be much better to *require* authors to register their works in order to enjoy the benefits of copyright protection. This would involve changing some international treaties, but some have suggested intriguing workarounds. For example, one could retain current copyright practices in order to comply with treaties, but limit statutory damages for infringement to \$1 unless the work was registered.

The U.S. Copyright Office maintains an online, searchable registry for those works that are registered there [3]. If it contained records of *all* U.S. copyrighted works, it would be far more useful. Even better, such a registry could contain the fee (if any) required for reproduction. This would make it much, much easier for digital material to be disseminated while, at the same time, respecting authors' rights.

True, there are existing institutions, such as the Copyright Clearance Center [1], that play this role for

certain sorts of content. However, there is still a considerable amount of material that is copyrighted but for which it is virtually impossible to find the rights holder.

For these cases, we need a legal safe harbor provision for orphan works—those works for which a rights holder cannot be identified. For example, if someone wants to copy a work and cannot find the rights holder in the registry, they can pay a nominal licensing fee in escrow. This money could be held for a period of time, in case a legitimate rights holder steps forth to claim it. If such a rights holder doesn't appear within an appropriate time period, the money could be used for other good purpose—for example, literacy programs or subsidizing deployment of information infrastructure.

To its credit, the U.S. Copyright Office has recently solicited comments on how best to handle "orphan works." Perhaps momentum is building to create some sort of standardized safe harbor policy.

If such a policy were in place, access to digitized information could be substantially enlarged. It would also be easier to experiment with other sorts of intellectual property rights such as those advocated by the Creative Commons, open archive projects, the GNU project, and so on.

Universal access to all the world's information is technologically possible now; the missing piece is the legal infrastructure that will provide the incentives to make such access economically viable.

The two legal reforms described here—a complete copyright registry for acquiring reproduction rights and a legal safe harbor for orphan works—would go a long way toward improving universal access to information. ■

REFERENCES

1. Copyright Clearance Center; www.copyright.com.
2. Lyman, P. and Varian, H.R. How much information 2003; www.sims.berkeley.edu/research/projects/how-much-info-2003/.
3. United States Copyright Office; www.copyright.gov/.
4. Wedgeworth, R.W. State of adult literacy 2004. ProLiteracy; www.proliteracy.org/downloads/StateofLit04.pdf.

HAL VARIAN (hal@sims.berkeley.edu) is a professor of business, economics, and information management at the University of California, Berkeley.