# Too big to close down: Websites need regulation like utilities

By Dov Greenbaum and Mark Gerstein

Google recently began the long process of shutting down its once popular software distribution platform, Google Code.

Why should you care?

Because this isn’t the first time a content provider has closed down a service, and it won’t be the last. Although we grudgingly acknowledge that our online information is being mined for salable nuggets, we may be less aware of the other deal-with-the-devil with free online storage and services: without hard copies, we are overly reliant on the supposedly enduring, albeit untested, viability and usability of our online host sites.

With untold petabytes of memorable photographs, critical e-mails, favorite music and important documents likely saved for posterity only online, the possible consequences from this unfounded reliance may be greater than previously imagined. Moreover, it’s not just the actual data that can get lost — it’s time: users also spend countless hours uploading and customizing workflows for each particular site.

And it gets worse: Vinton Cerf, often referred to as a father of the Internet, recently foreshadowed a new and imminent Dark Ages because of the predicted huge data loss through obsolescence of our current data storage systems. However, Cerf’s is not just a prediction of some far off future. The axiom that the “Internet never forgets” is a misnomer. The Internet loses boat loads of data every time “[www.your\_new\_favorite\_website\_for\_storing\_all\_your\_important\_data.com](http://www.your_new_favorite_website_for_storing_all_your_important_data.com" \t "_blank)” closes down; remember all those flashing gifs and god-awful midi music of Yahoo’s Geocities? Neither does the Internet.

But while Cerf’s proposal, Digital Vellum, may be useful for large institutional libraries and long-term archival storage, the general public needs a simple to employ solution so that when the next Friendster or Kodak Gallery goes dark, or the next Digg is re-engineered and customized workflows no longer work, users are guaranteed both sufficient notification and data portability before all is lost.

While Google ought to be commended in how it allows for the easy exporting of Google data through its Takeout system, not everyone is Google. The Internet isn’t really an option anymore. Like a public utility, people need to be able to rely on the continued viability of their service providers and their stuff. With so much of society’s permanent record represented only online, these sites have an implicit obligation to protect the permanence of that record. However, we can’t rely on profit-motivated companies to maintain what are, in all fairness, collectively mostly junk, but to each of us individually, irreplaceable memories.

As such, we propose a regime wherein once a website or service hits a predetermined number of active users, it becomes too big to simply close down without legal repercussions. Fundamentally, the site takes on some of the character of a public utility and is perceived as such by regulatory bodies. Standards and their enforcement could be through industry committees, or a statutorily imposed duty-bound relationships. Effectively, content hosting sites with sufficient active users would have to employ open and/or standard formats, for example as are commonly used in e-mail systems and online calendars. Further, these sites would have to either provide for user-friendly data import and export tools, or allow for such tools to be developed by interested third parties.

In the event that the service provider can no longer maintain the data, e.g., for financial reasons or business-related concerns, the users must be provided with sufficient notification of the imminent loss of their data. Moreover, in general, users should always be provided with the ability to move their data to another provider or service, should they choose, and the service provider must actively help users move their data once notification is provided that the data will be lost.

While many companies already provide some of these services, particularly in the examples listed above, it is important that we enforce these protections universally. Establishing universal portability standards are important, and everyone will benefit from the knowledge that they can always move their data elsewhere. Furthermore, becoming a “public” standard and utility is very flattering to the service provider — even though it comes at a cost. While, the public should also be educated to help themselves and create regular backups, new, naïve and/or unsophisticated users are the ones most likely to be harmed when a website goes unexpectedly dark.

By developing and enforcing broad comprehensive protections, we will be able to protect even those that haven’t taken otherwise necessary precautions to protect their data.

The closing down of your favorite website is inevitable, likely as a result of corporate and technological growth and change, but it is doesn’t have to be overly disrupting.

Dov Greenbaum is director of the Zvi Meitar Institute for Legal Implications of Emerging Technologies at the Radzyner Law School, Interdisciplinary Center, Herzliya, Israel and a professor in Molecular Biophysics and Biochemistry at Yale University. Mark Gerstein is the A. L. Williams Professor of Biomedical Informatics at Yale University.