

Watch this: corralling wild bits

[Terrence A. Brooks](#)
The Information School
University of Washington
Seattle, WA 98195

[Here's the gist! Answer this question! The bits of this column are residing in your machine right now. What will you do with them? Toss them? Archive them? Manage them? How much discretionary time and idle effort have you got to devote to the task, whatever your choice?]

When Bits Are Wild Intellectual Capital Escapes Watch institutions retain bits in an attempt to recoup costs of producing intellectual capital. Watch individuals preserve bits to organize memory and create identity. Watch governments mine bit repositories to generate intelligence.

Driving metaphors:

- Our *shoeboxes* full of old photographs prove the value of archiving - **MyLifeBits**
- Punching holes in *stovepipes* of information leads to discovery - **Total Information Awareness**
- Information *living inside* asset management systems can be repurposed- **DSpace**

But Retaining Information Can Cut Both Ways: *Maybe it's good:* Memex was proposed more than 50 years ago by Vannevar Bush: everything would be copied into the desktop archive and linked together semantically. Fifty years later we have the Semantic Web. *Maybe it's bad:* Oval Office tapes capturing all conversation in the Nixon White House proved too damaging and a mysterious 18 1/2 minute erasure occurred in the tape of June 20, 1972.

Announcement: MyLifeBits

[MyLifeBits](#) is a Microsoft project for storing all of one's digital media, including documents, images, sounds, videos, phone calls, television programs. You can keep your life stored as bits. You could do a Google search on your life.

Supposing one did keep virtually everything--would there be any value to it? Well, there is an existence proof of value. The following exist in abundance: **shoeboxes** full of photos, photo albums & framed photos, home movies / videos, old bundles of letters bookshelves and filing cabinets. (Gemmell, et al, 2002)

For further information:

Bass, M.J. (2002) [DSpace - a sustainable solution for institutional digital asset services - spanning the information asset value chain: ingest, manage, preserve, disseminate](#). [version 2002-03-01]
<http://dspace.org/technology/architecture.pdf> (22 Mar. 2003)

Ebbinghaus, H. (1964) *Memory: a Contribution to experimental psychology*. New York, NY: Dover Publications.

Gemmell, J., Bell, G., Lueder, R., Drucker, S. & Wong, C. (2002) [MyLifeBits: Fulfilling the Memex Vision](#). [Paper delivered at] *ACM Multimedia '02*, December 1-6, 2002 Juan Les Pins, France.
<http://research.microsoft.com/~jgemmell/pubs/MyLifeBitsMM02.pdf> (22 Mar. 2003)

Hackos, JoAnn T. (2002) *Content Management for Dynamic Web Delivery*. New York, NY: John Wiley.

The Rockley Group, Inc. (2003) [Managing enterprise content: a unified content strategy. White paper, revised Mar. 14, 2003](#). Markham, Ontario: The Rockley Group Inc.
<http://www.rockley.com/articles/The%20Rockley%20Group%20-%20ECM%20UCS%20Whitepaper%20-%20revised.pdf> (22 Mar. 2003)

Reinhardt, Ad. (1975) *Art-as-art: the selected writings of Ad Reinhardt*. Edited and with an introduction by Barbara Rose. Berkeley, CA: University of California Press.

Reality check! Do you have the discretionary time to archive this column somewhere? Give it a name that is meaningful to you? Attach keywords to your copy? And if you don't, do you trust some automatic indexing algorithm to determine this column's meaning for you? And if you relent and just stream content into a digital repository, haven't you created the equivalent of a digital shoebox? What could happen if the police did a Google search on your life, or even worse, an estranged former spouse? a stalker? Or, God forbid, a telemarketer?

Announcement: DSpace

[DSpace](#) a digital repository project at MIT has attracted the cooperation of six major research universities: Columbia University, Cornell University, Ohio State University, and the Universities of Rochester, Toronto, and Washington. There is the promise of capturing faculty intellectual production, sharing it and repurposing it.

DSpace uses the rhetoric of content management: "Databases are information silos, but a distributed environment..."

means that historic silo-oriented views of information **living 'inside'** an asset management system will not be sufficient. Assets will be used by many parties, each with different world views, for many purposes. Systems and services in the future will flow around existing information assets, and standards-based mechanisms for access to those assets will need to be a part of the web infrastructure. (Bass, et al, 2002)

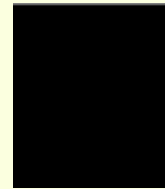
Sub-rosa economic argument: Did you know that the State of Washington is approximately \$2 billion in debt? Did you know that the University of Washington pays me a salary to write academic articles? In effect, the State of Washington pays me to write columns like this one; can you believe it? I write academic articles and then send them off to journals. The University of Washington then pays for a subscription to the academic journal (paper version) or leases access (to the proprietary database version) to gain access to my article? Did the tax payers of the State of Washington pay for that intellectual product *twice*? Will my Legislators in Olympia, WA recognize the economic opportunity of capturing the University's bits and vending them? Is DSpace is the precursor of systems that will capture the intellectual production of my university, catalogue it for retrieval and make it available for repurposing? **Question:** What will be the role of academic journals like this one when universities capture and control their bits?

Reality check! The beta version I used had a laborious cataloguing process. Someone has to do the traditional librarian's job of cataloguing the item. My personal commitment to sharing information is maintaining an online vita with HTML copies of my papers. As a busy scholar, I don't have time for this. Filling out the form below is for cataloguers.

Defense Advanced Research Projects Agency (2003) [TIA FAQs. DARPA's Information Awareness Office \(IAO\) and Total Information Awareness \(TIA\) Program. Frequency asked questions:](#) http://www.darpa.mil/iao/TIA_FAQs.pdf (22 Mar. 2003)

The First Entry In Your Electronic ShoeBox

Here is my homage to the black paintings of Ad Reinhardt (American Abstract painter, 1913-1967). Enter this into your collection, but first determine what it's about.



[Help for the bewildered:](#) Who is Ad Reinhardt? Why would he create black paintings? What could a black painting possible mean? Etc?

[Here is a screen shot of the entry form.](#) Do you have time to fill out this form?

Announcement: TIA

[*Total Information Awareness*](#) is a U.S. government initiative to gather and mine information for finding terrorists. The ambition is to coordinate all the information that the government possesses.

The collaborative reasoning and decision-support technologies will solve existing coordination problems by enabling analysts from one agency to collaborate effectively with analysts in other agencies. A major challenge to terrorist detection today is the inability to quickly search, correlate, and share data from databases maintained legally by our intelligence, counterintelligence and law enforcement agencies. The collaborative reasoning and decision-support technologies will punch holes in these 'stovepipes.' (TIA FAQs)

Reality check! Many Americans don't regard governments as innately benign. A safeguard for democracy is distributing power widely and not concentrating too much power (information) in a few hands. *Historical note:* I'm writing this as the United States is attacking Iraq. Do I really want to give this government more of my personal information than they already have? **Answer:** No!

Ok, I'm ready to corral my bits

Question: What's your bursting configuration?

Content re-use is the strategy of using existing content components to create new documents. If you intend on re-using information, you either construct it in a technology designed for re-use, e.g.: XML - Extensible Markup Language, or you burst apart existing documents into their constituents parts, e.g.: images, text, etc.

With component-management systems, topics developed using structured content units with appropriate tags (either SGML, XML, or style tags) are burst apart or shredded for storage. Individual content units are stored as separate objects in the component-management system rather than stored as whole documents or modules. Consequently, the individual content units are available for reuse in different assemblies. (Hackos, 2002: 79)

Streaming bits into an archive is not content management and limits the possibility of re-use. What is the level of granularity of the DSpace entry form above? Do I really have sufficient

discretionary time to ponder the granularity of the content I would stream into MyLifeBits?

Question: What happened to erasure and anonymity in the Digital Age?

When you use Internet Explorer a history cache is created, images are saved. Your HTTP header information announces who you are, what you asked from the server and all of this activity is tracked in server logs. It is easy to sniff the byte stream to your computer and monitor your wireless communications. Effectively we have lost our ability of erasure in the digital age. Every digital thing I touch leaves a track back to me.

Forgetting goes unappreciated in the digital age:

In 1885, researcher Thomas Ebbinghaus did a study on people's ability to retain information. He called the results the "Curve of Forgetfulness."

Ebbinghaus found that a person forgets 75% of what he or she has learned in the previous week. After three weeks, he/she forgets 90%. After four weeks, he/she forgets 95%. The Ebbinghaus study is an illustration of retentiveness; the brain retains information it considers important to the individual and "forgets" information not deemed relevant.

[Frequency = Success](#)

Forgetting becomes impossible when your delete button disappears:

Note that for all but video, the delete operation may well become obsolete: the user's time for the delete operation will be more costly than the storage to keep the item. (Gemmell, et al.)

Question: What's the difference between spam and poetry?

Digital content grows exponentially and now even *machines* can generate e-mails aimed at my in-box. Have we confused "content" with bit streams? In the legacy paper technology world, information was scarce and I hoarded my paper sources of information, but I'm victimized by information in the digital world. To set up a Hot Mail account is to become a spam target. I definitely would not want to archive this in MyLifeBits. I know people who have multiple e-mail addresses to fend off unwanted information. Using "Reply to All" becomes an anti-social act.

The only factor becoming scarce in a world of abundance is human attention.

[Kevin Kelly, Maxims for the Network Economy](#)

Is there such a thing as information that has no meaning? bits without content? Can you create a painting that is designed to

defeat viewing?

Art is art. Everything else is everything else.
Ad Reinhardt

Information is information. Everything else is
everything else.
Brooks' First Corollary of Spam

Accept the challenge in the side bar and enter one of Reinhardt's
black paintings in your digital shoebox.

Date: March 2003

How to cite this paper:

Brooks, T.A. (2003) "Watch this: corralling wild bits". *Information Research*, **8**(3) paper no. TB0304 [Available at
<http://InformationR.net/ir/8-3/TB0304.html>]
© the author, 2003 Updated: 22 March 2003

[Contents](#)


[Web Counter](#)

[Home](#)
