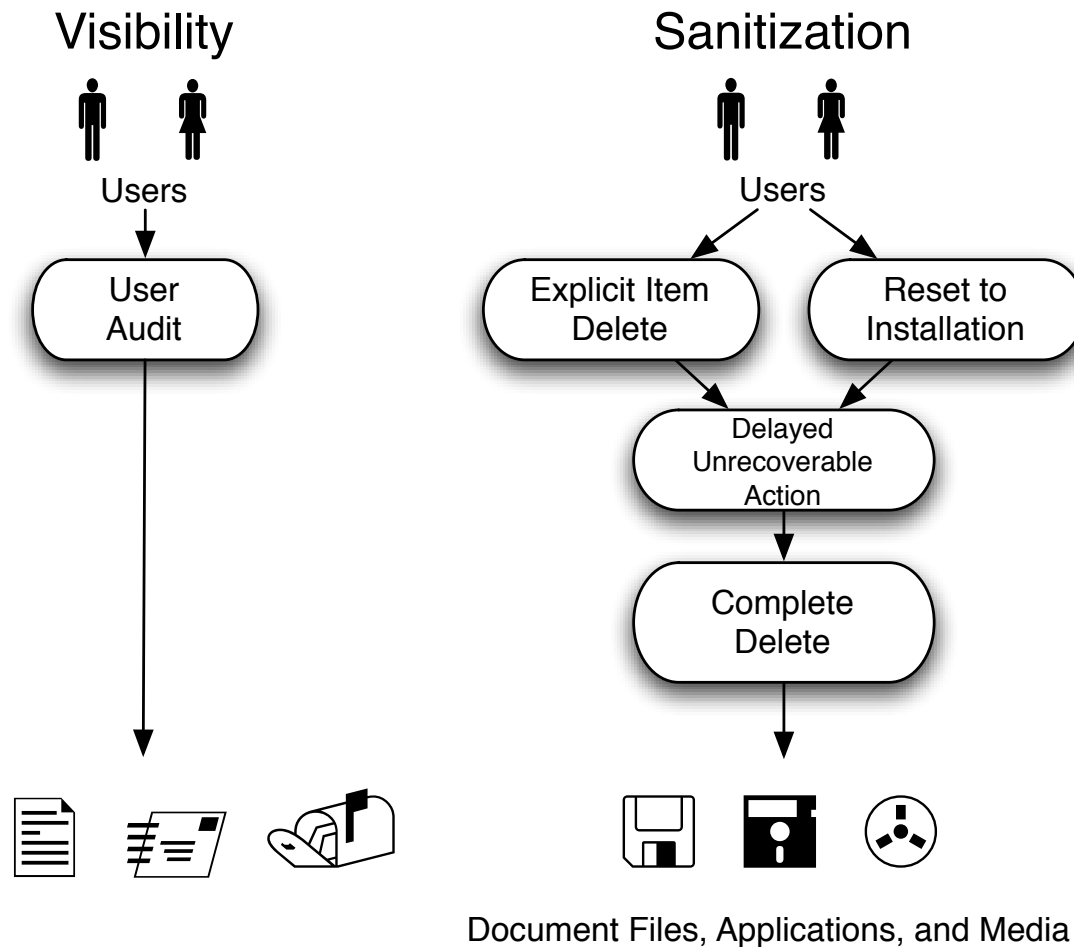
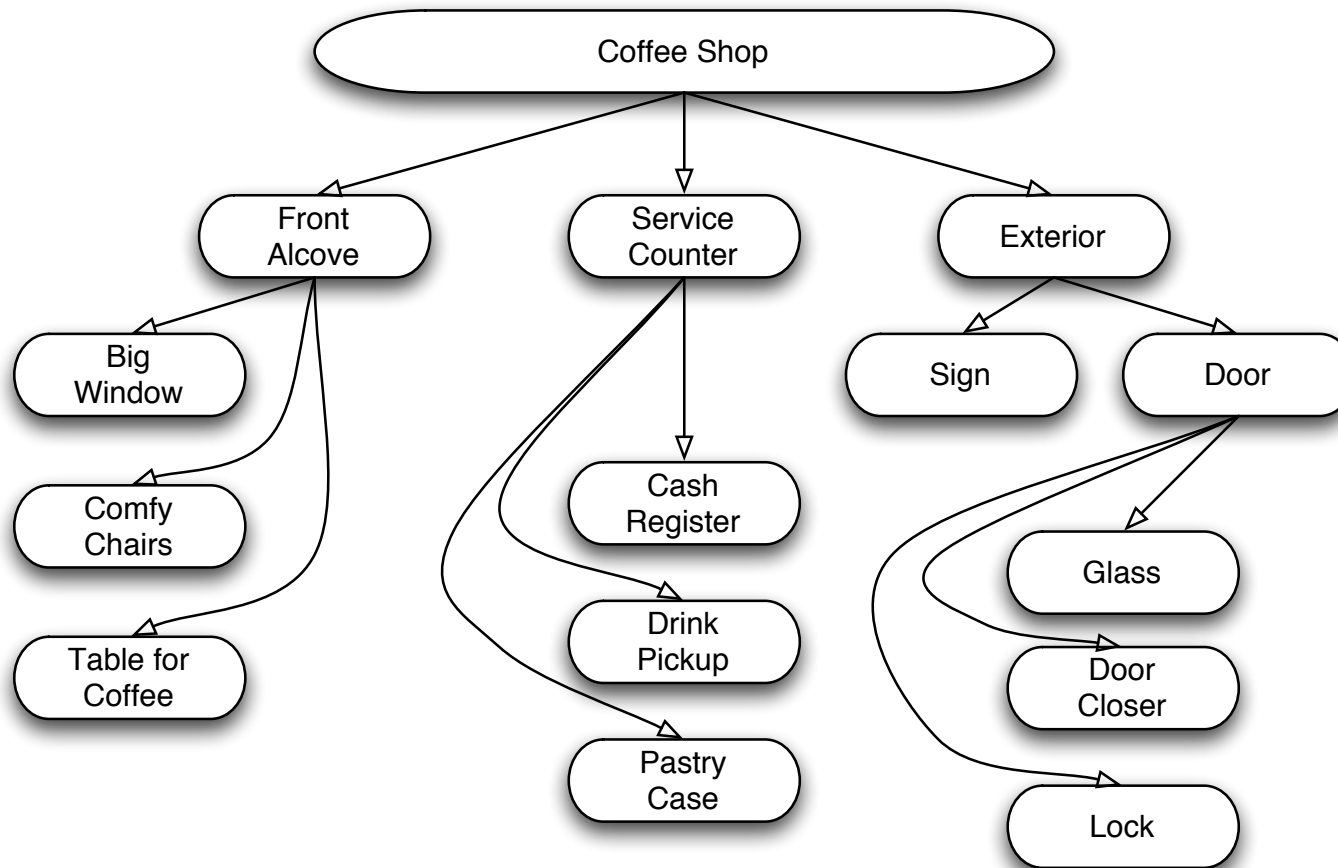


“Complete Delete” and Other Patterns for Information Eradication



Simson L. Garfinkel
Center for Research on Computation and Society
Harvard University
October 11, 2005

A pattern is a recurring solution to a standard problem.

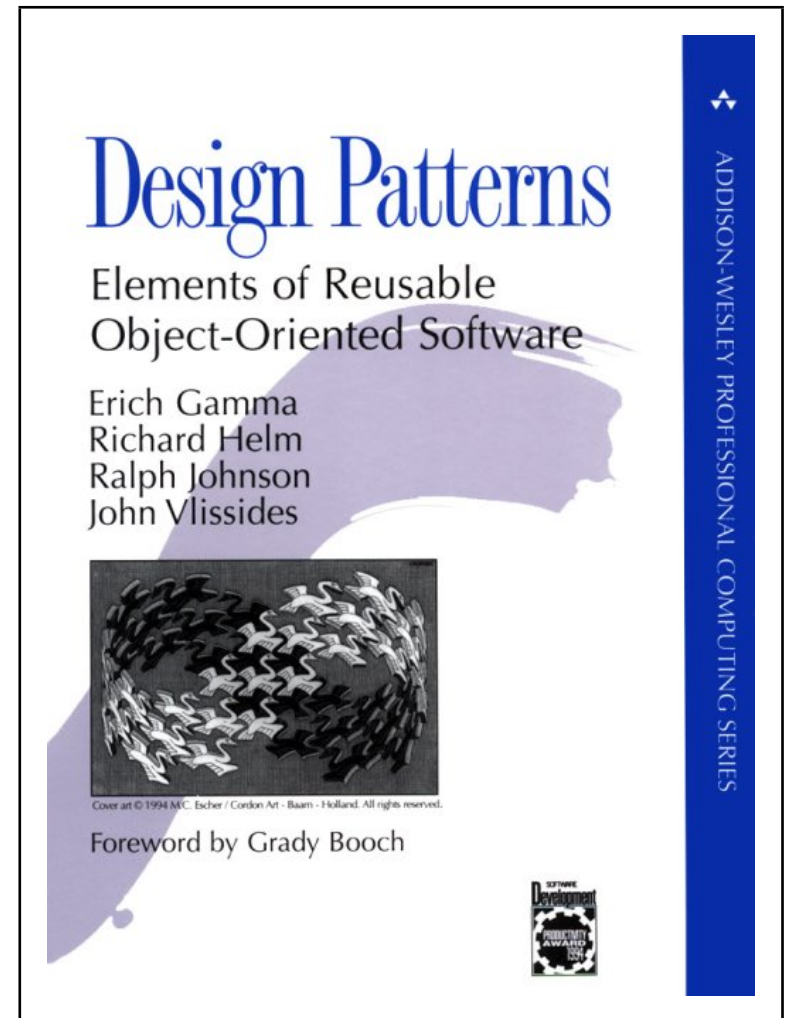


Patterns and “pattern languages” introduced by Architect Christopher Alexander in the 1970s.

Object Oriented Design adopted patterns in the 1990s.
Johnson *et al.*, [OOPSLA 91]; Coad [CACM, 92]; “Gang of four” [95]

Patterns help us to:

- reuse successful practices
- reason about what’s done and why
- document abstractions other than algorithms and data structures.
[Schmidt *et al.*, 1996]



**Patterns encapsulate knowledge and understanding,
making it easier to teach and deploy solutions.**

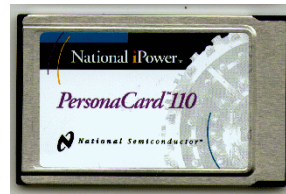
It has long been recognized that end-user security and usability are at odds in modern computer systems.



Username: simsong
Password: •••••



ACCESS DENIED



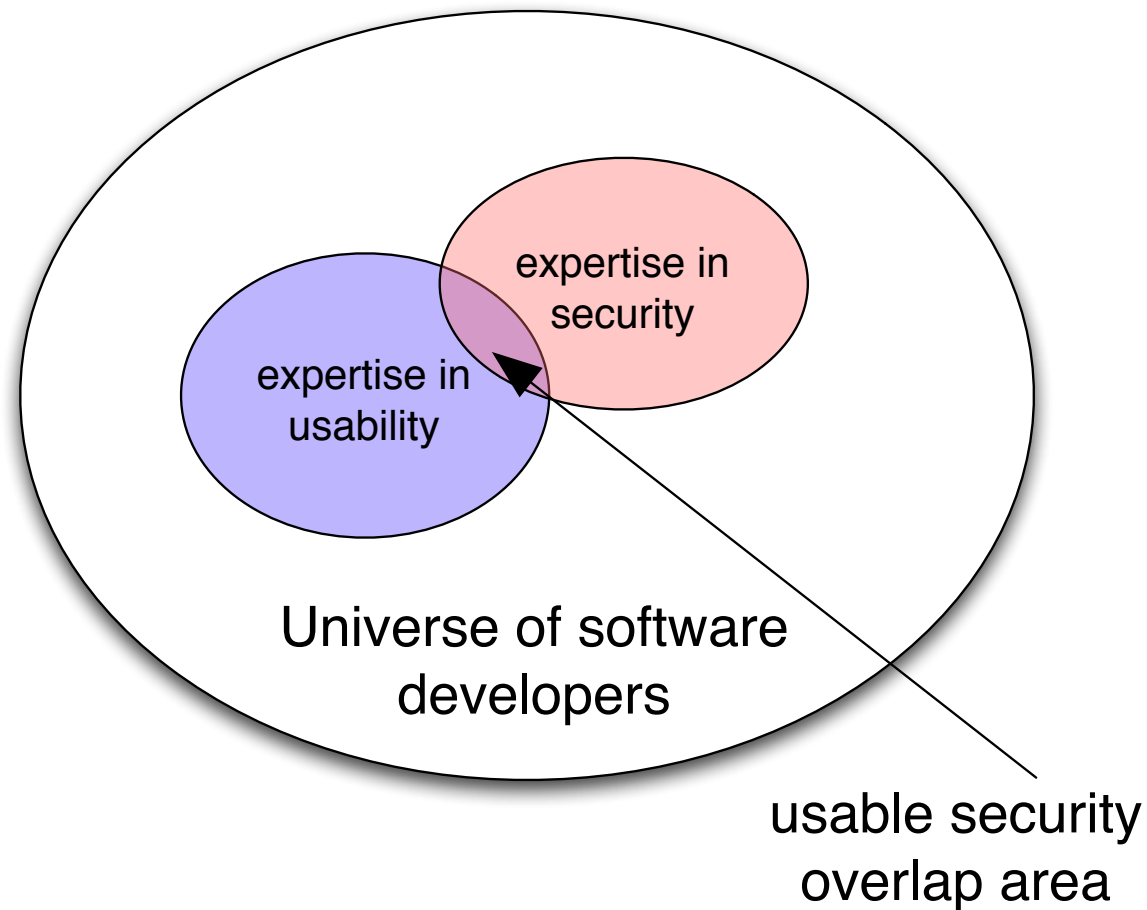
The need to align end-user security and usability is recognized as a priority for both computing and the nation.

- CRA 2003 “Grand Challenge”
- PITAC 2005 “priority”
- Special publications
[IEEE S&P 2004] [O’Reilly 2005]
- CHI 2005; SOUPS 2005



HCI-SEC is the emerging field that seeks to align Human Computer Interfaces with Security.

The root of the conflict: security and usability *must both be applied from the beginning*—but they are *different skills*.



Today computer security has many “principles,” “best practices” and “techniques.”

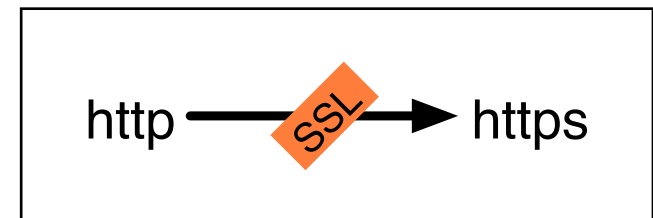
- Biometric authentication



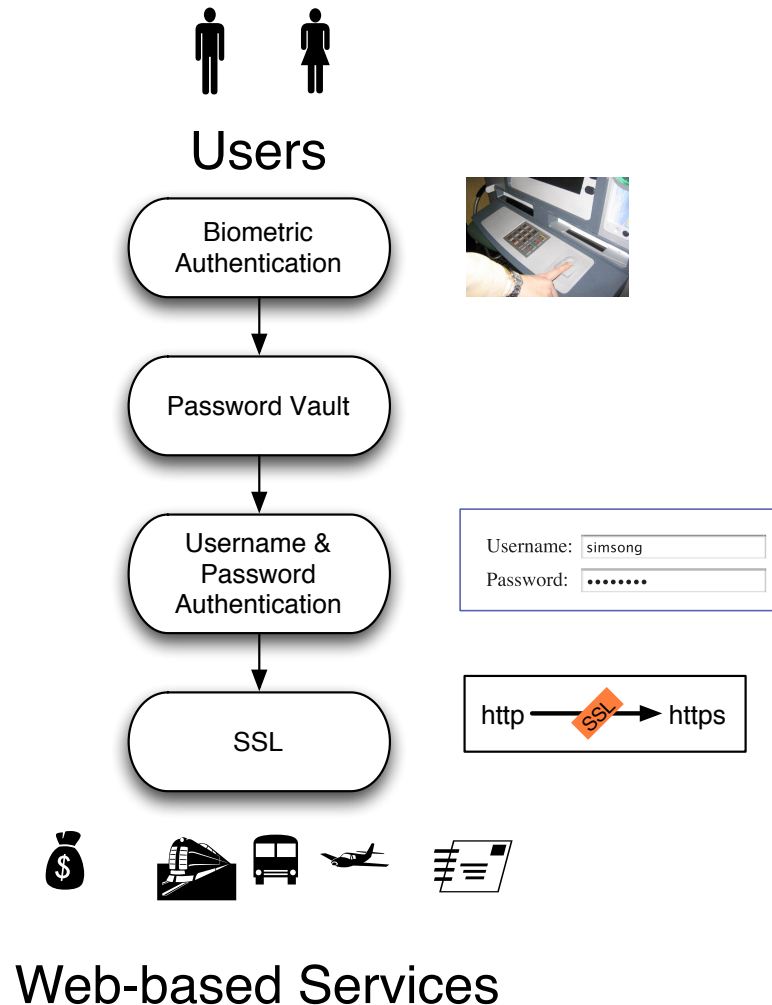
- The password field

Username:	<input type="text" value="simsong"/>
Password:	<input type="password" value="....."/>

- Wrapping plaintext protocols with SSL



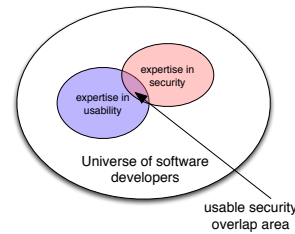
With patterns, we can decompose the problems and their solutions.



Patterns make it easier to communicate solutions to students, implementors, and management.

This talk presents set of patterns for destroying information.

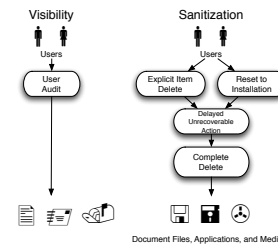
1. Introduction to patterns. ✓



2. The Drives Study



3. Patterns for sanitization.



4. Applying these patterns.



The Sanitization Problem: Confidential information is left behind after it is no longer needed.

Data discovered on second-hand hard drives is an obvious case.

- Woman in Nevada bought a used PC with pharmacy records [Markoff 97]
- Paul McCartney's bank records sold by his bank [Leyden 04]
- Pennsylvania sold PCs with “thousands of files” on state employees [Villano 02]



**Between January 1999 and April 2002,
I acquired 236 hard drives on the secondary market.**



[Garfinkel & Shelat 03] established the scale of the problem.

We found:

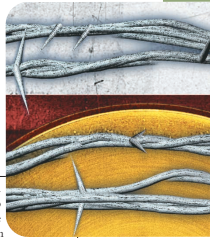
- Thousands of credit card numbers (many disks)
- Financial records
- Medical information
- Trade secrets
- Highly personal information

We did not determine if this was a *usability* problem or an *education* problem.

Data Forensics

Remembrance of Data Passed: A Study of Disk Sanitization Practices

Many discarded hard drives contain information that is both confidential and recoverable, as the authors' own experiment shows. The availability of this information is little publicized, but awareness of it will surely spread.



A fundamental goal of information security is to design computer systems that prevent the unauthorized disclosure of confidential information. There are many ways to assure this information privacy. One of the oldest and most common techniques is physical isolation: keeping confidential data on computers that only authorized individuals can access. Most single-user personal computers, for example, contain information that is confidential to that user.

Computer systems used by people with varying authorization levels typically employ authentication, access control lists, and a privileged operating system to maintain information privacy. Much of information security research over the past 30 years has centered on improving authentication techniques and developing methods to assure that computer systems properly implement these access control rules.

Cryptography is another tool that can assure information privacy. Users can encrypt data as it is sent and decrypt it at the intended destination, using, for example, the secure sockets layer (SSL) encryption protocol. They can also encrypt information stored on a computer's disk so that the information is accessible only to those with the appropriate decryption key. Cryptographic file systems¹⁻³ ask for a password or key on startup, after which they automatically encrypt data as it's written to a disk and decrypt the data as it's read; if the disk is stolen, the data will be inaccessible to the thief. Yet despite the availability of cryptographic file systems, the general public rarely seems to use them.

Absent a cryptographic file system, confidential information is readily accessible when owners improperly reformat their disk drives. In August 2002, for example, the United States Veterans Administration Medical Center in Indianapolis retired 139 computers. Some of these systems were donated to schools, while others were sold on the open market, and at least three ended up in a thrift shop where a journalist purchased them. Unfortunately, the VA neglected to *sanitize* the computer's hard drives—that is, it failed to remove the drives' confidential information. Many of the computers were later found to contain sensitive medical information, including the names of veterans with AIDS and mental health problems. The new owners also found 44 credit card numbers that the Indianapolis facility used.⁴

The VA fiasco is just one of many celebrated cases in which an organization entrusted with confidential information neglected to properly sanitize hard disks before disposing of computers. Other cases include:

- In the spring of 2002, the Pennsylvania Department of Labor and Industry sold a collection of computers to local resellers. The computers contained "thousands of files of information about state employees" that the department had failed to remove.⁵
- In August 2001, Dovebid auctioned off more than 100 computers from the San Francisco office of the Viant consulting firm. The hard drives contained confidential client information that Viant had failed to remove.⁶
- A Purdue University student purchased a used Macintosh computer at the school's surplus equipment exchange facility, only to discover that the computer's hard drive contained a FileMaker database containing the names and demographic information for more than 100 applicants to the school's Entomology Department.
- In August 1998, one of the authors purchased 10 used computer systems from a local computer store. The

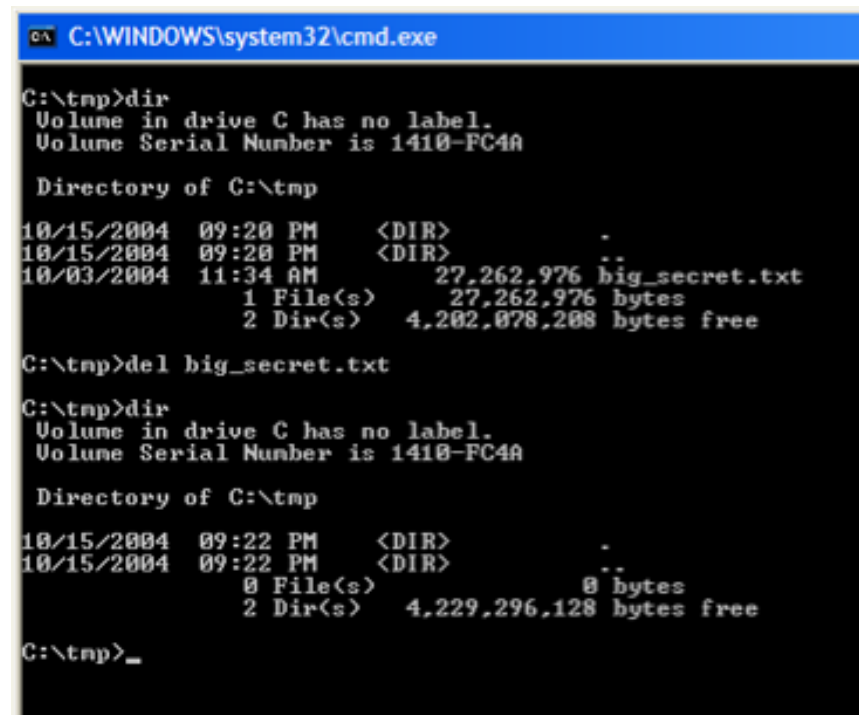
SIMSON L. GARFINKEL AND ABHI SHELAT
Massachusetts Institute of Technology

PUBLISHED BY THE IEEE COMPUTER SOCIETY ■ 1540-7993/03/17-00 © 2003 IEEE ■ IEEE SECURITY & PRIVACY 17

Evidence for the usability problem: Computers *lie* when users delete data.

DEL removes file names

—but not file contents.



```
C:\WINDOWS\system32\cmd.exe

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

Directory of C:\tmp

10/15/2004  09:20 PM    <DIR>          .
10/15/2004  09:20 PM    <DIR>          ..
10/03/2004  11:34 AM                27,262,976 big_secret.txt
               1 File(s)                27,262,976 bytes
               2 Dir(s)          4,202,078,208 bytes free

C:\tmp>del big_secret.txt

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

Directory of C:\tmp

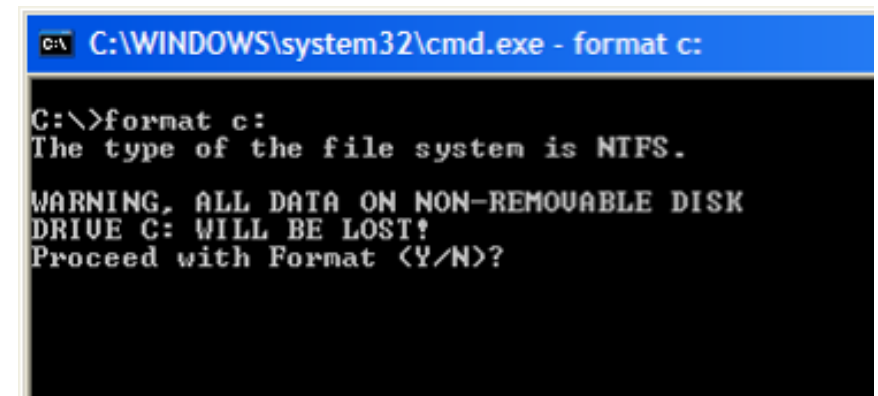
10/15/2004  09:22 PM    <DIR>          .
10/15/2004  09:22 PM    <DIR>          ..
               0 File(s)                  0 bytes
               2 Dir(s)          4,229,296,128 bytes free

C:\tmp>_
```

FORMAT claims

“ALL DATA ... WILL BE LOST”

—but it’s not.



```
C:\WINDOWS\system32\cmd.exe - format c:

C:\>format c:
The type of the file system is NTFS.

WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

This violates the principle of “Psychological Acceptability.”
[SS '75]

Psychological Acceptability: “... avoid a fundamental mismatch between software capabilities and the user’s mental models.”

“We all sincerely believed that when we ... pressed the button 'delete' that it was gone forever.

Wow, were we wrong.”

— Oliver North, 1987



Oliver North had a mismatched mental model.

Evidence for an educational problem: There is a huge secondary market for used disk drives.



- Re-used within organizations
- Given to charities
- Sold on eBay

All Categories [Save this search](#)

350 items found for **hard drives**

Sort by items: [ending first](#) | [newly listed](#) | [lowest priced](#) | [highest priced](#)

Picture	Item Title	Price	Bids	Time Left
	Lot of hard and floppy drives	\$5.50	2	14m
	Lot of hard and floppy drives	\$5.50	2	22m
	Lot of hard and floppy drives	\$5.50	2	25m
	Lot of 2 hard drives IDE	\$8.00	12	29m
	3.2 gig Hard Drives	\$180.00	-	59m
	(5) 1.2 hard drives & (15) 10/100 network	\$15.00	1	1h 00m
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	\$16.00	6	1h 25m
	IDE HARD DRIVES (3)	\$6.50	6	1h 46m
	LOT OF 5 Hard Drives! 3.2 Gig Western Digital	\$120.00 \$124.95 7 Apr 10	-	1h 50m
	QTY 3... IDE Hard Drives 2.5 Gg	\$10.50	5	2h 02m
	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	\$30.00	4	2h 03m
	QTY 3... IDE Hard Drives 1.0 Gg	\$9.99	1	2h 04m
	Western Digital 850 meg IDE Hard Drives dutch	\$6.00	1	2h 57m
	WINDOWS	\$6.00	-	3h 18m

People could just be discarding disk drives without thinking about the consequences.

To be effective, patterns should address the root cause of the problem.

Usability Problem:

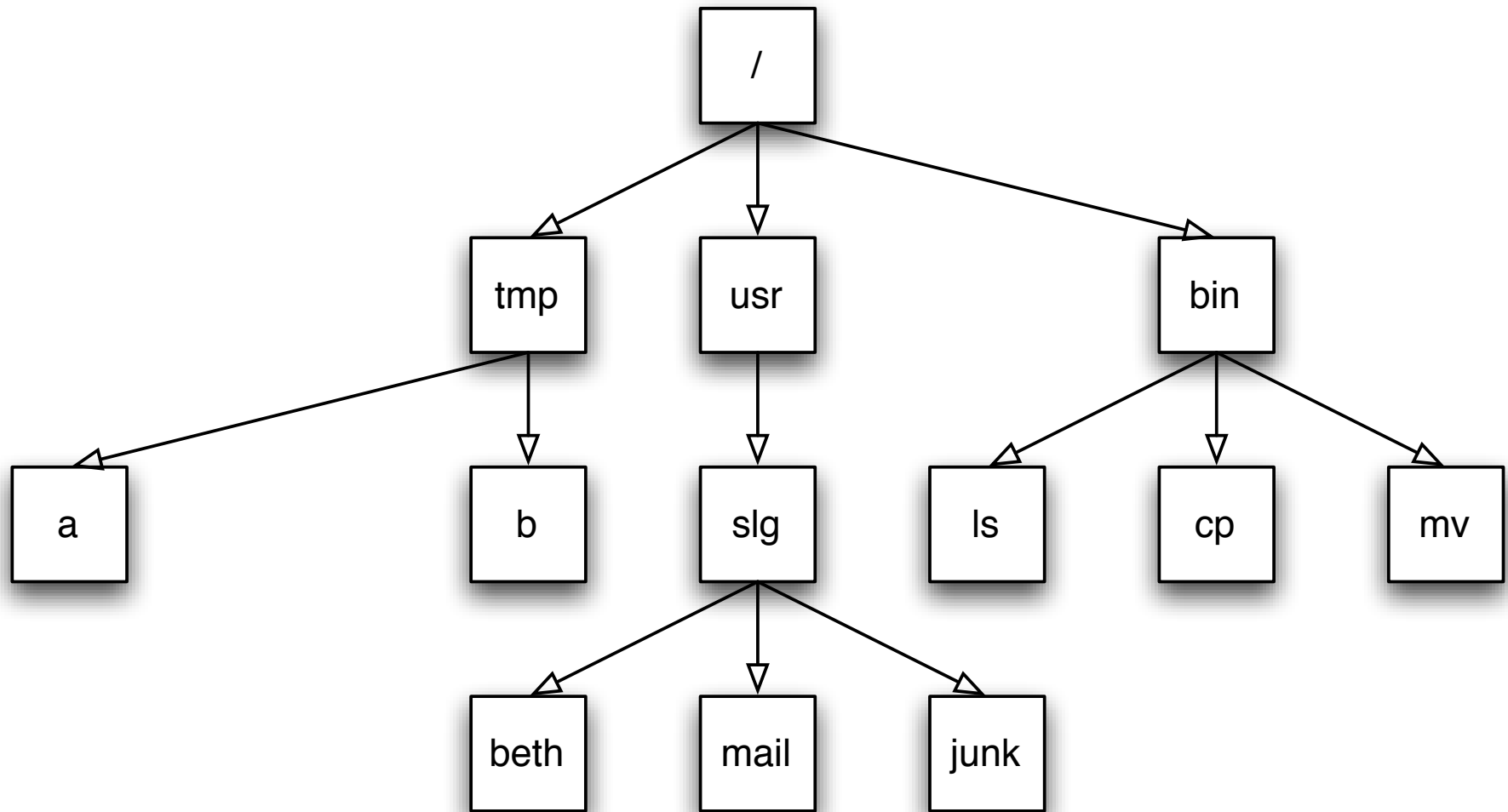
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data.
[Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

Education Problem:

- Add training to the interface.
[Whitten 04]
- Regulatory requirements.
[FTC 05, SEC 05]
- Legal liability.

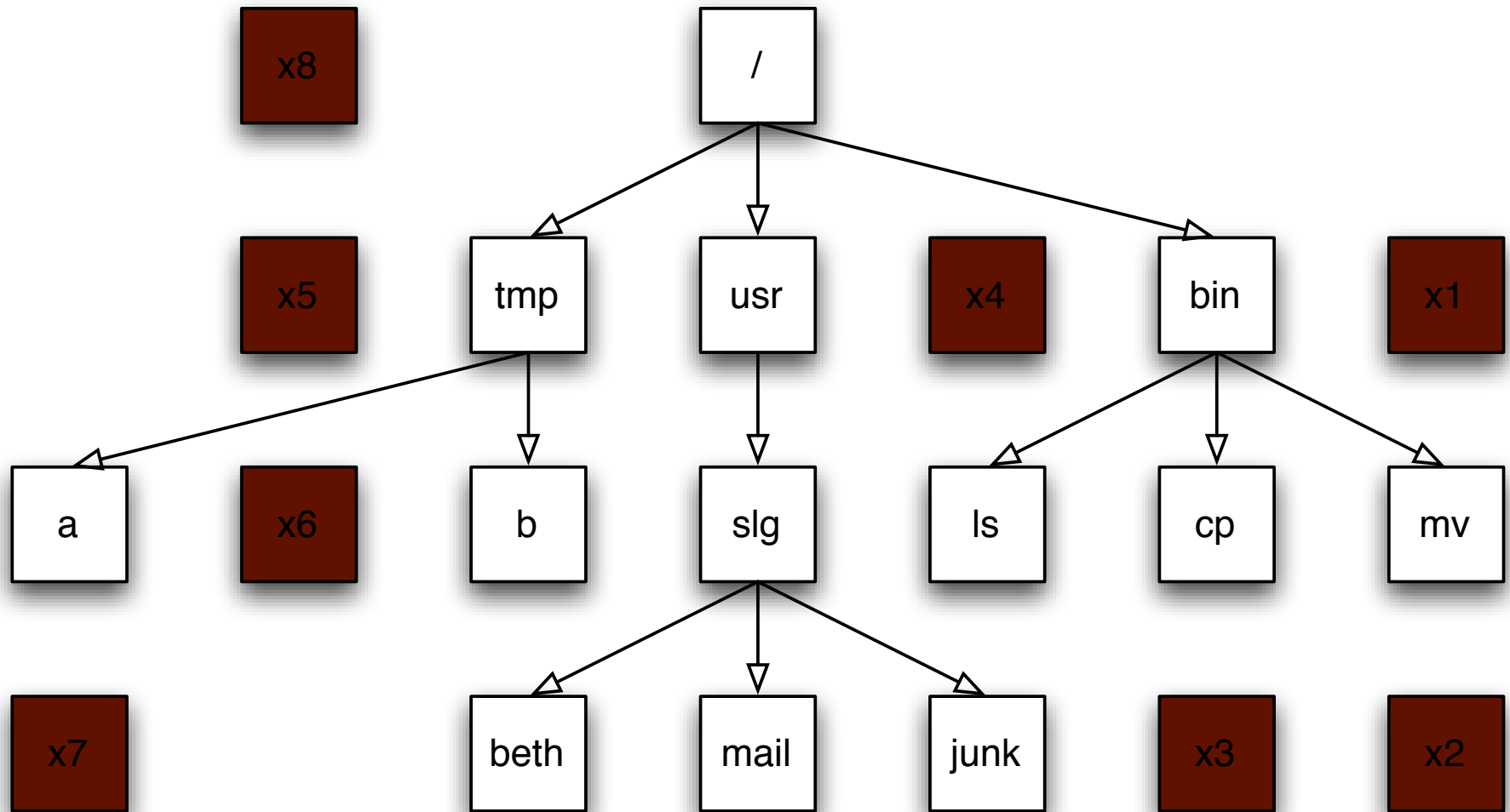
To determine the root cause, I looked *on the drives* and *contacted the data subjects*.

Data on a hard drive is arranged in blocks.



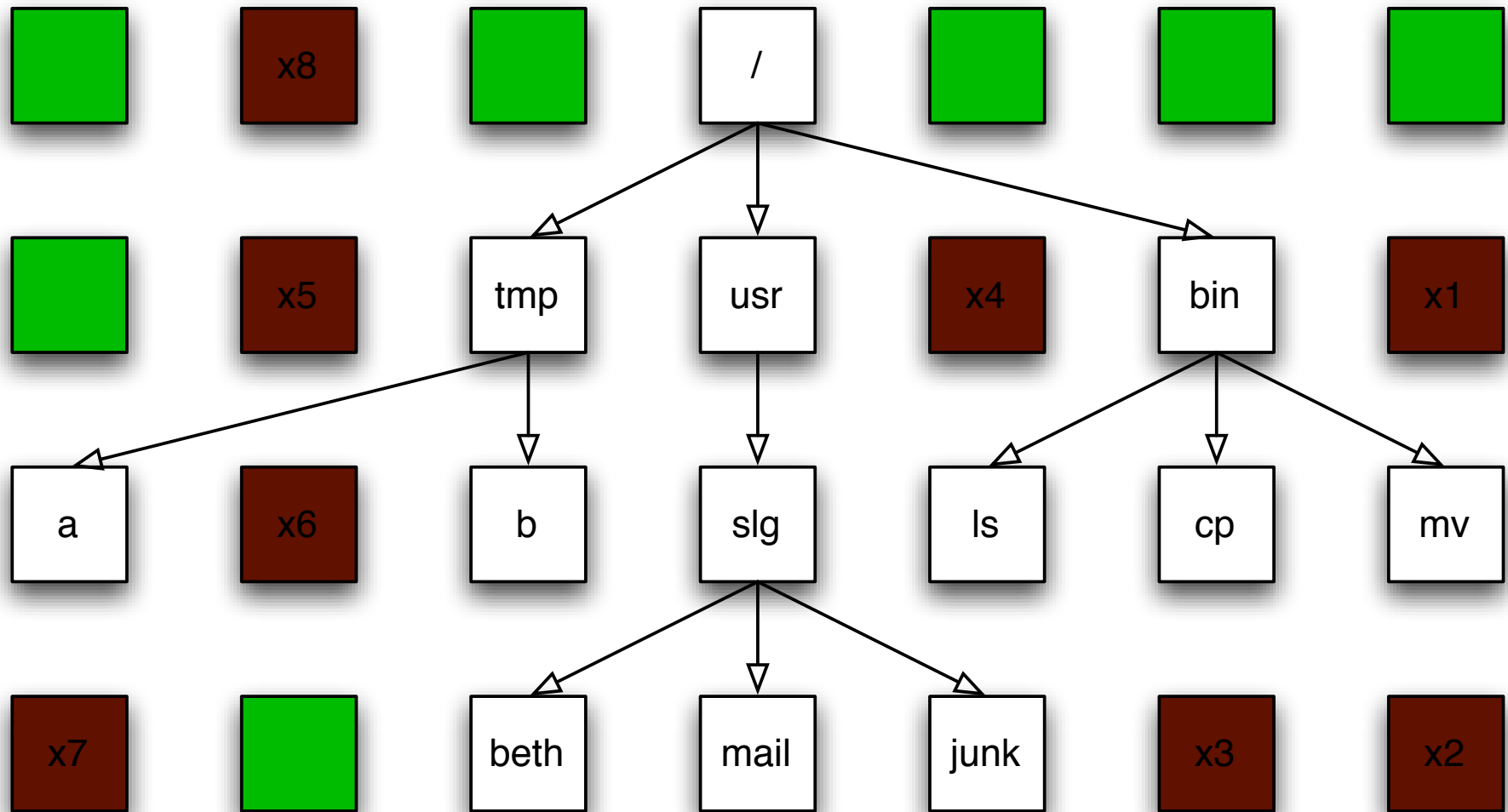
The white blocks indicate directories and files that are visible to the user.

Data on a hard drive is arranged in blocks.



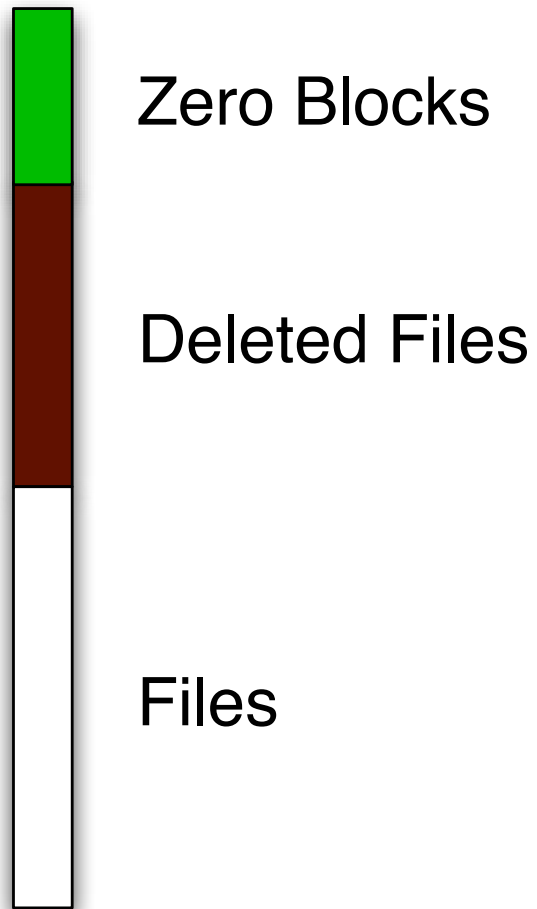
The brown blocks indicate files that were deleted.

Data on a hard drive is arranged in blocks.

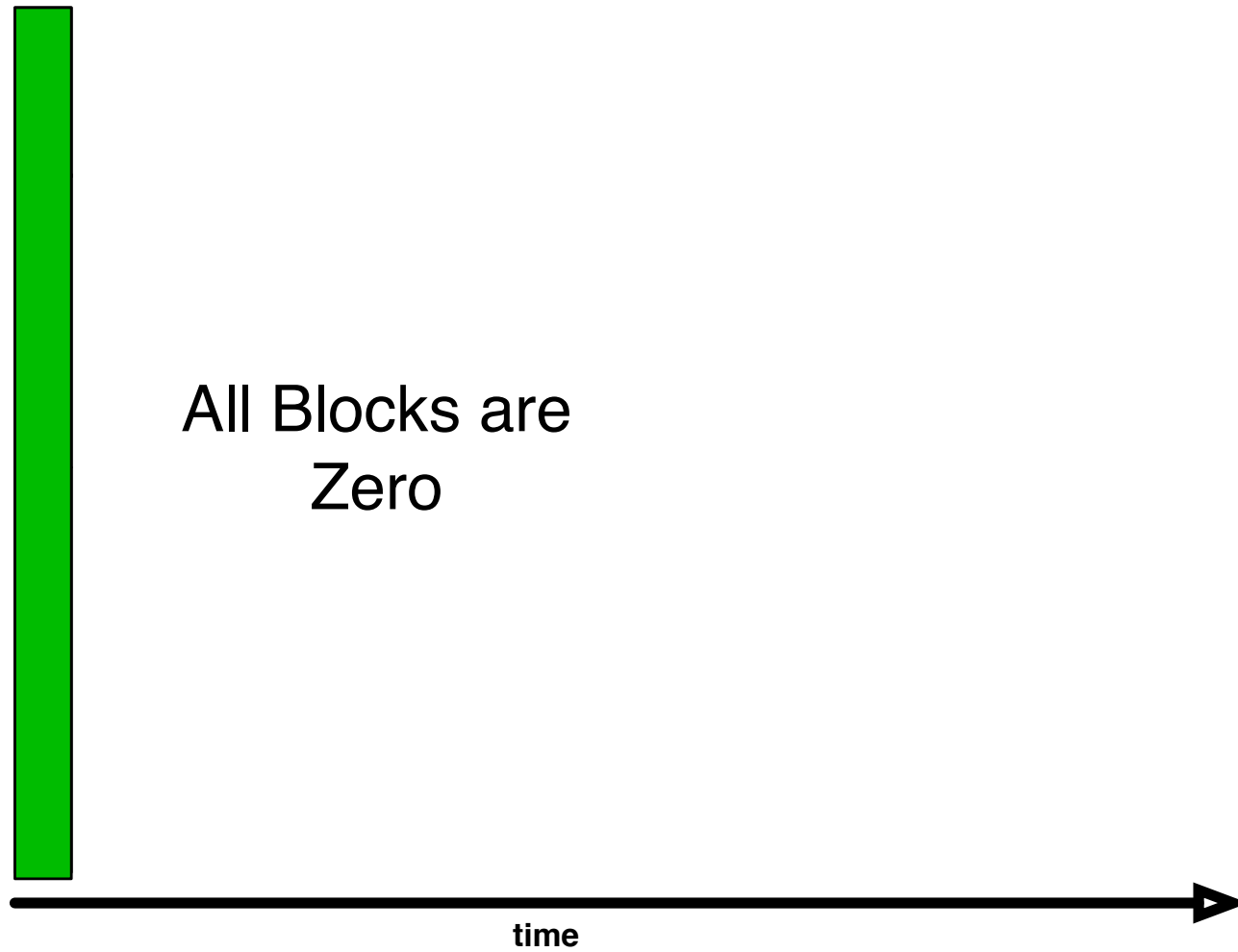


The green blocks indicate blocks that were never used (or that were wiped clean).

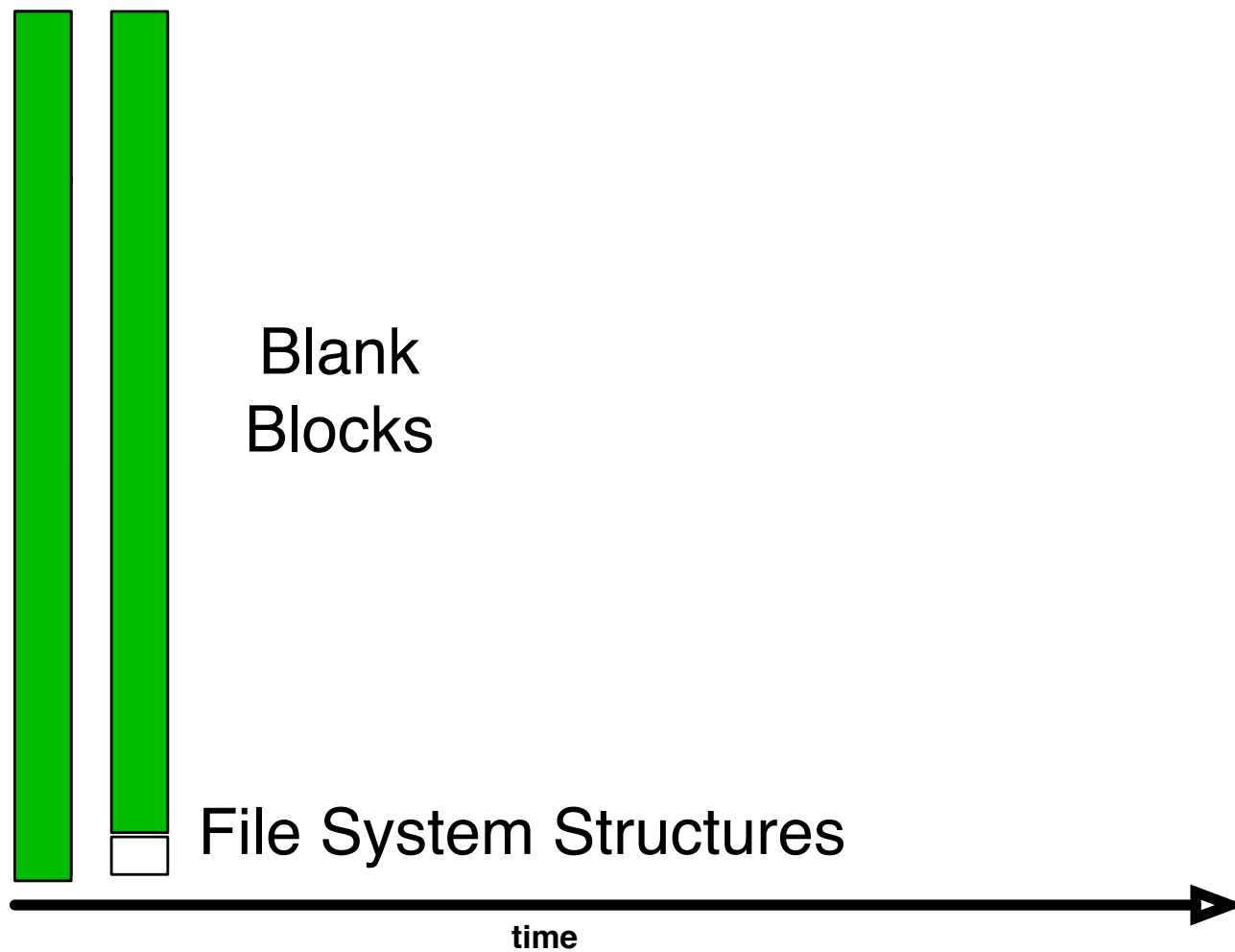
Stack the disk blocks:



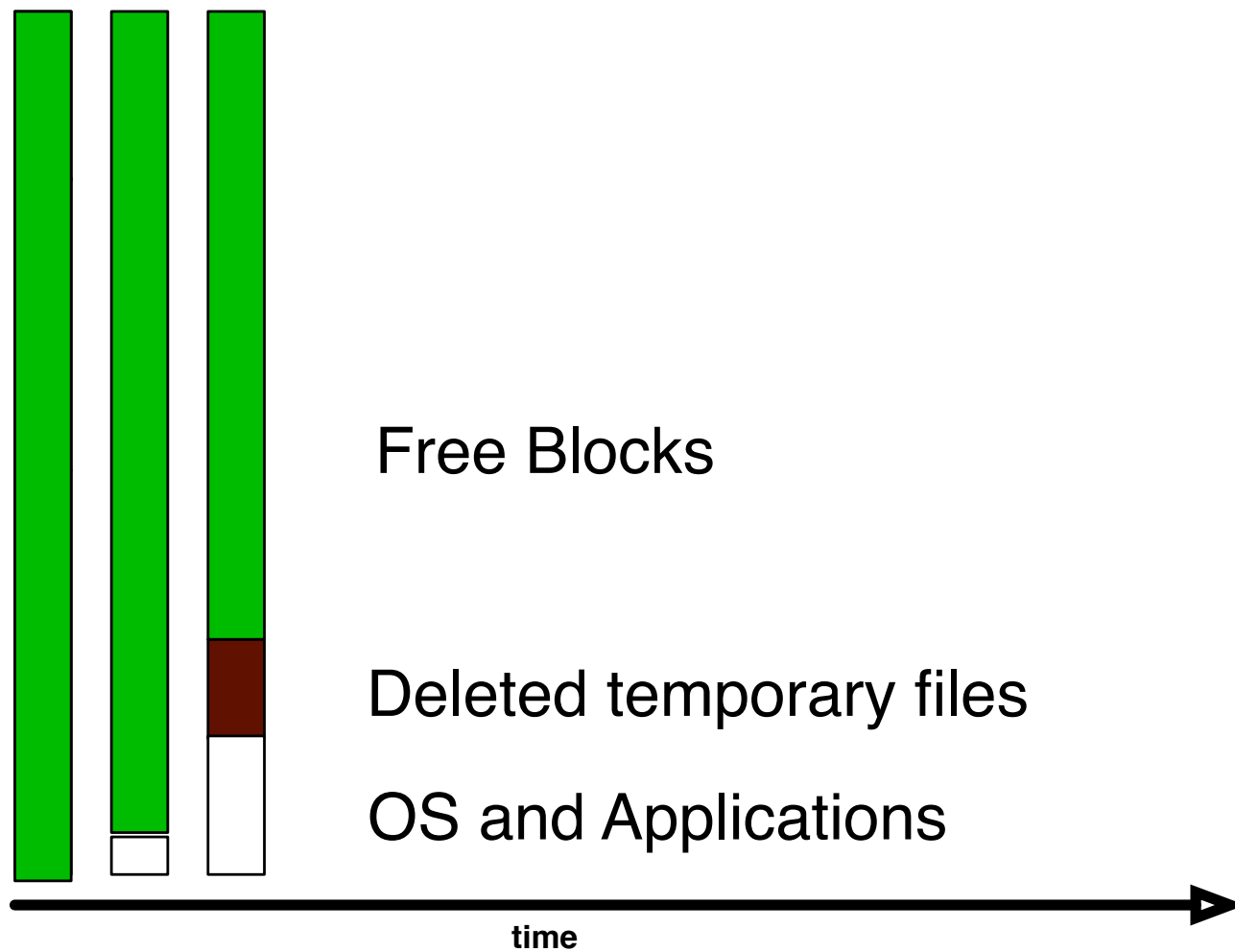
NO DATA: The disk is factory fresh.



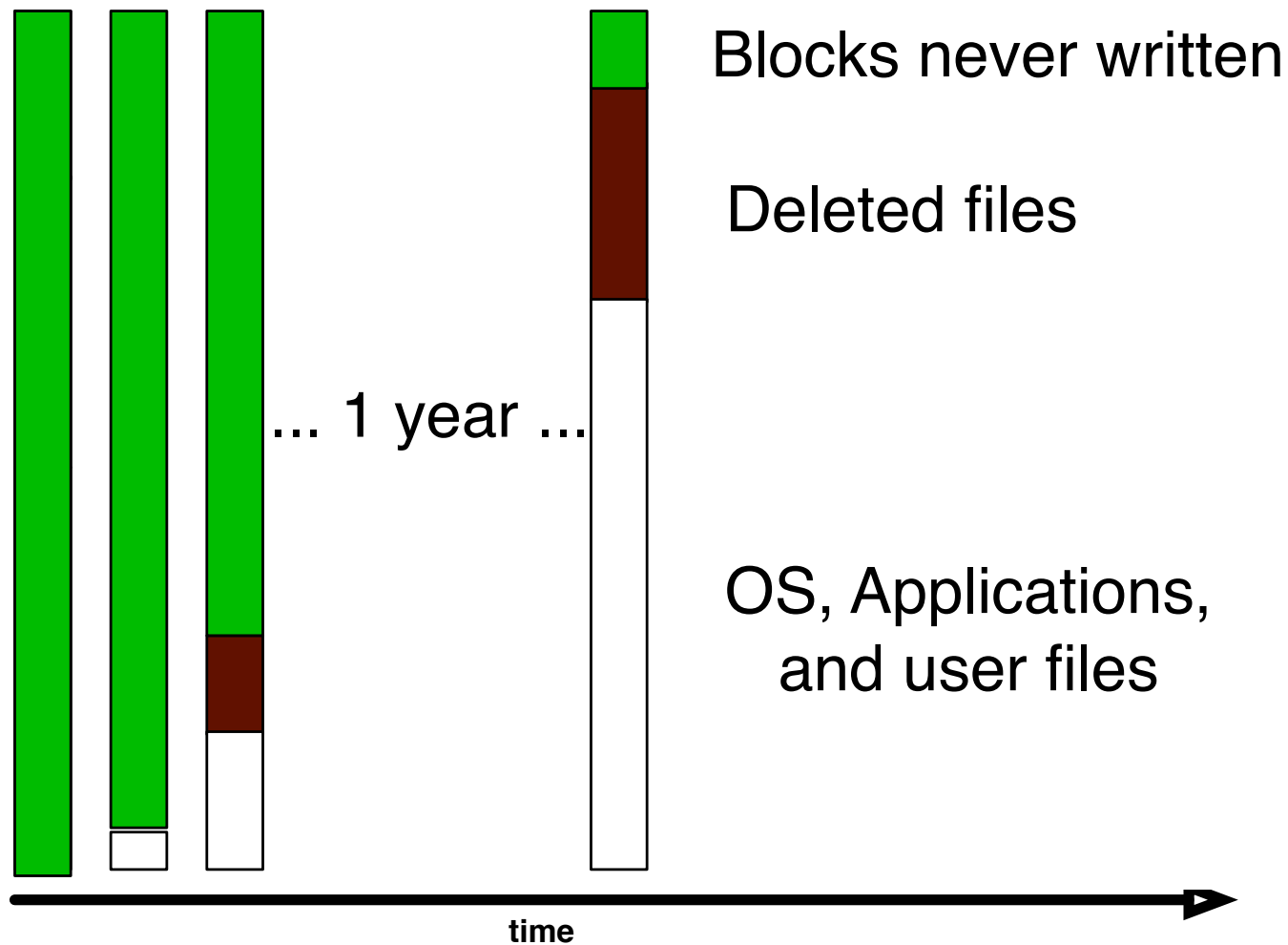
FORMATTED: The disk has an empty file system



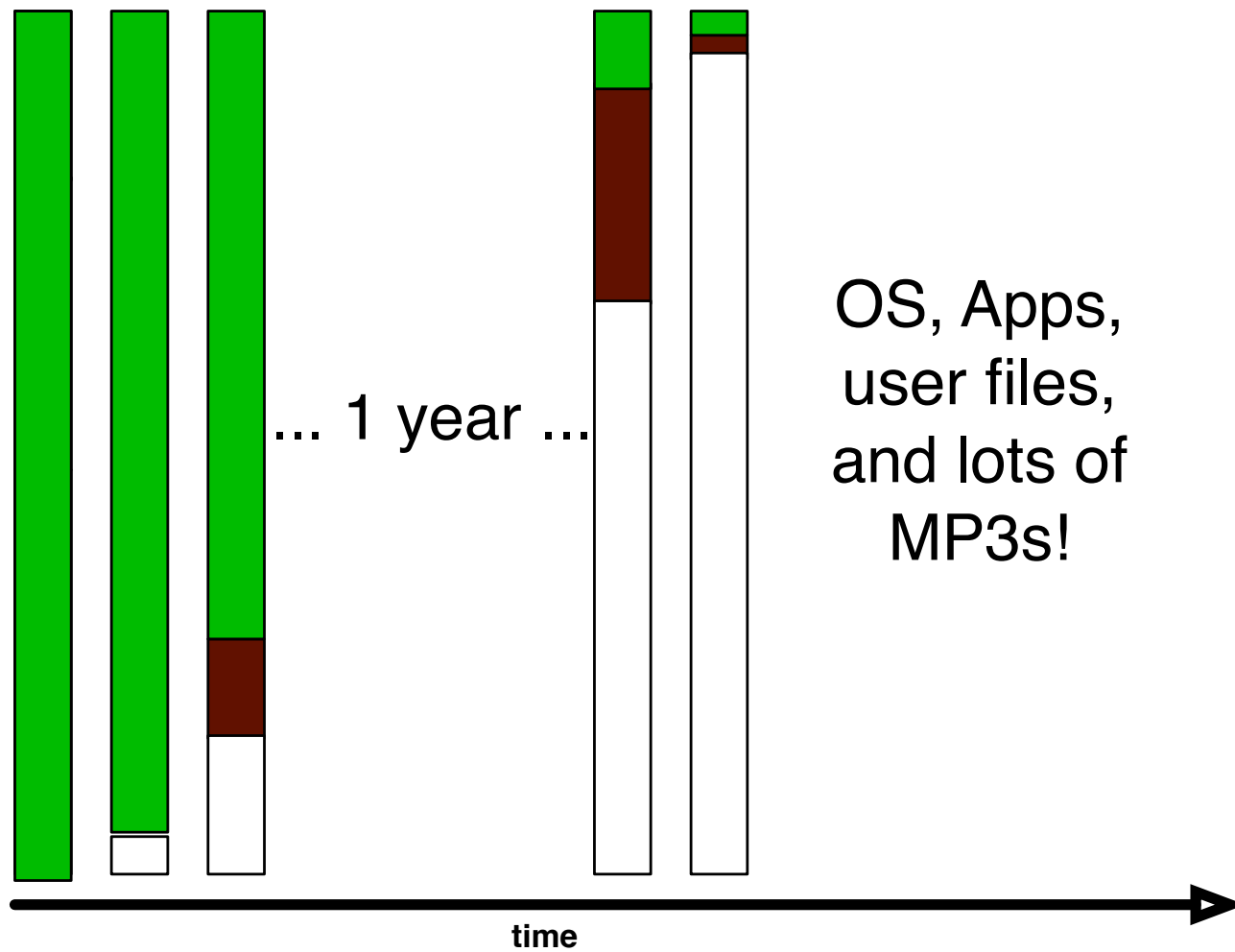
AFTER OS INSTALL: Temp. files have been deleted



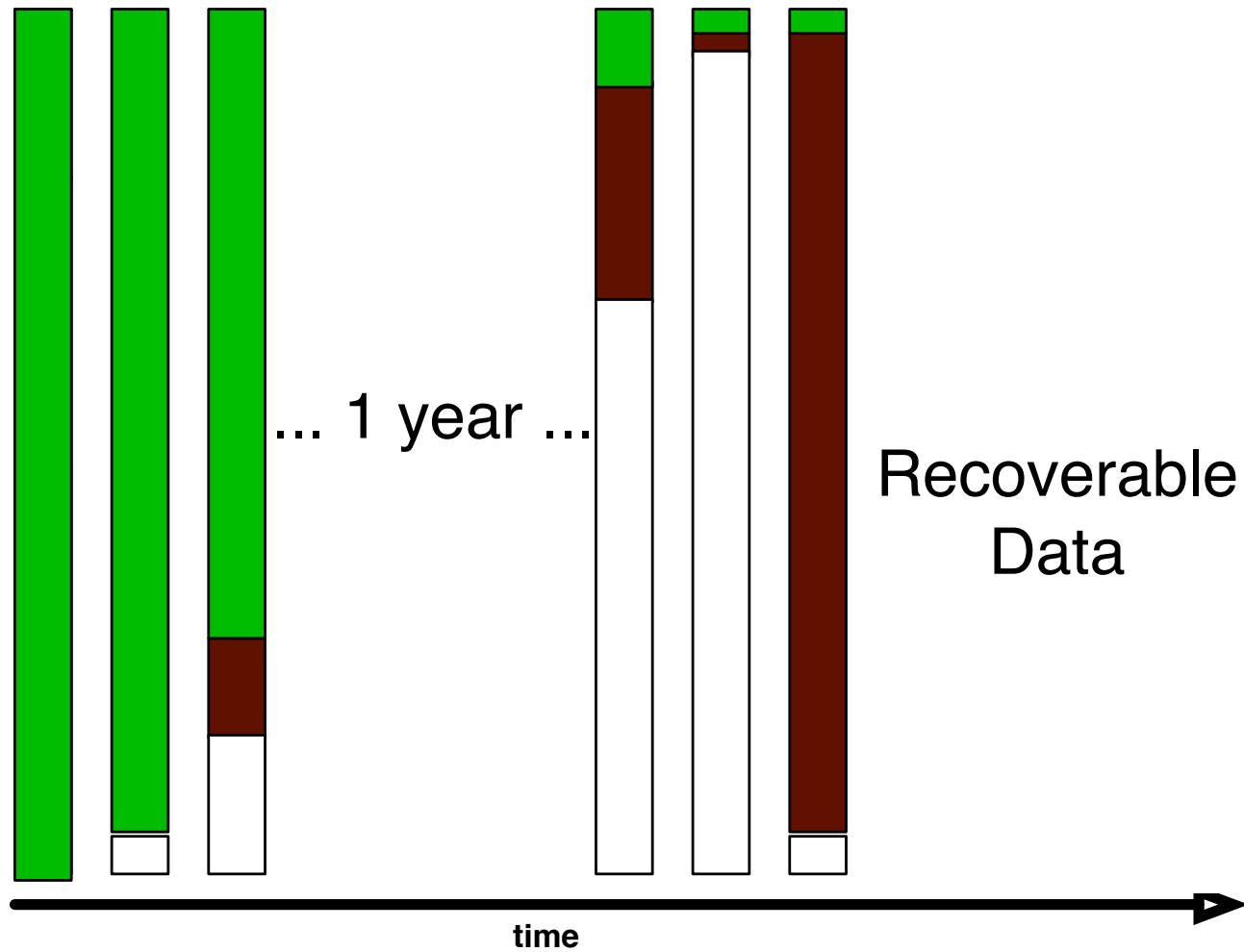
AFTER A YEAR OF SERVICE



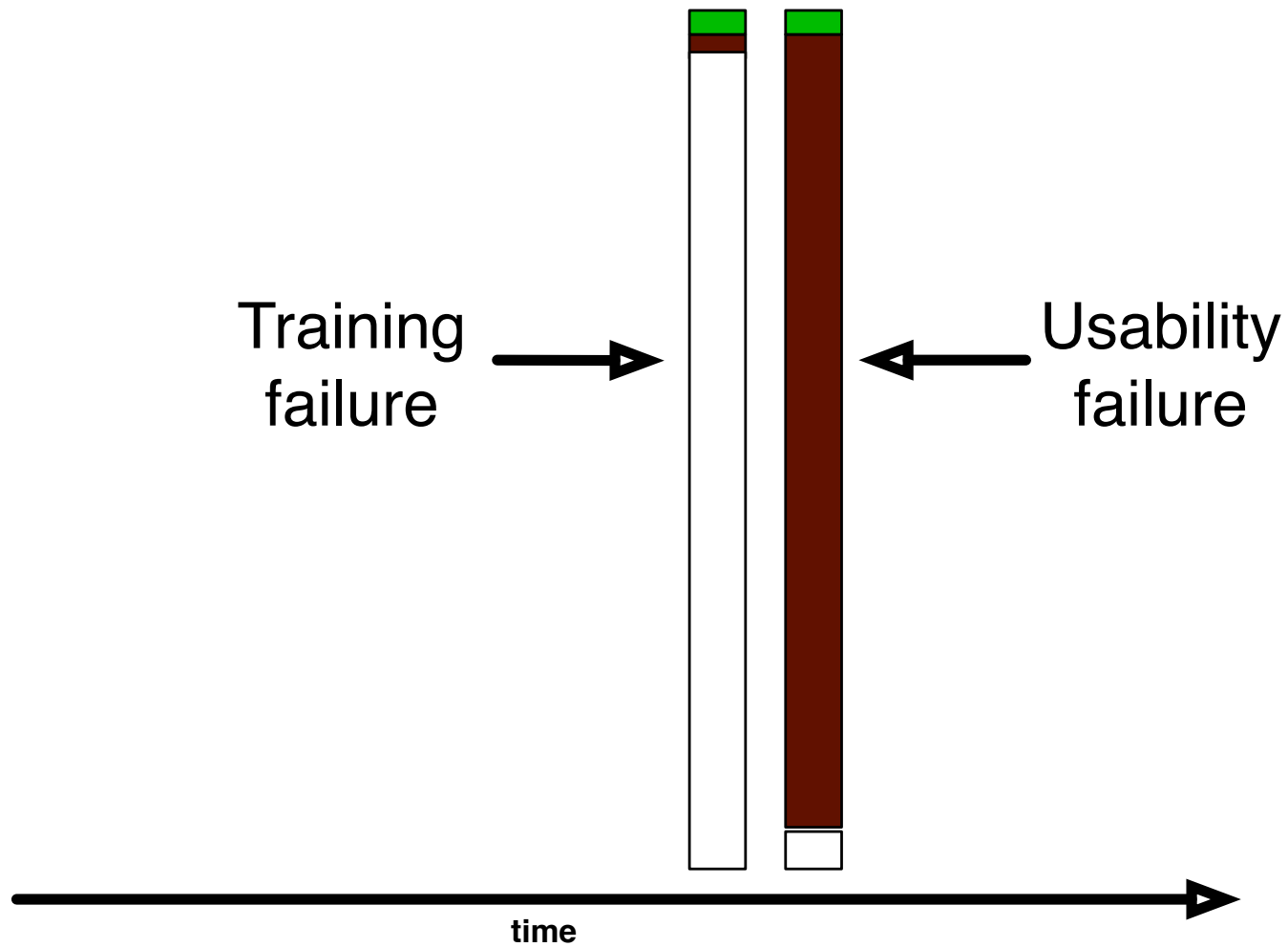
DISK NEARLY FULL!



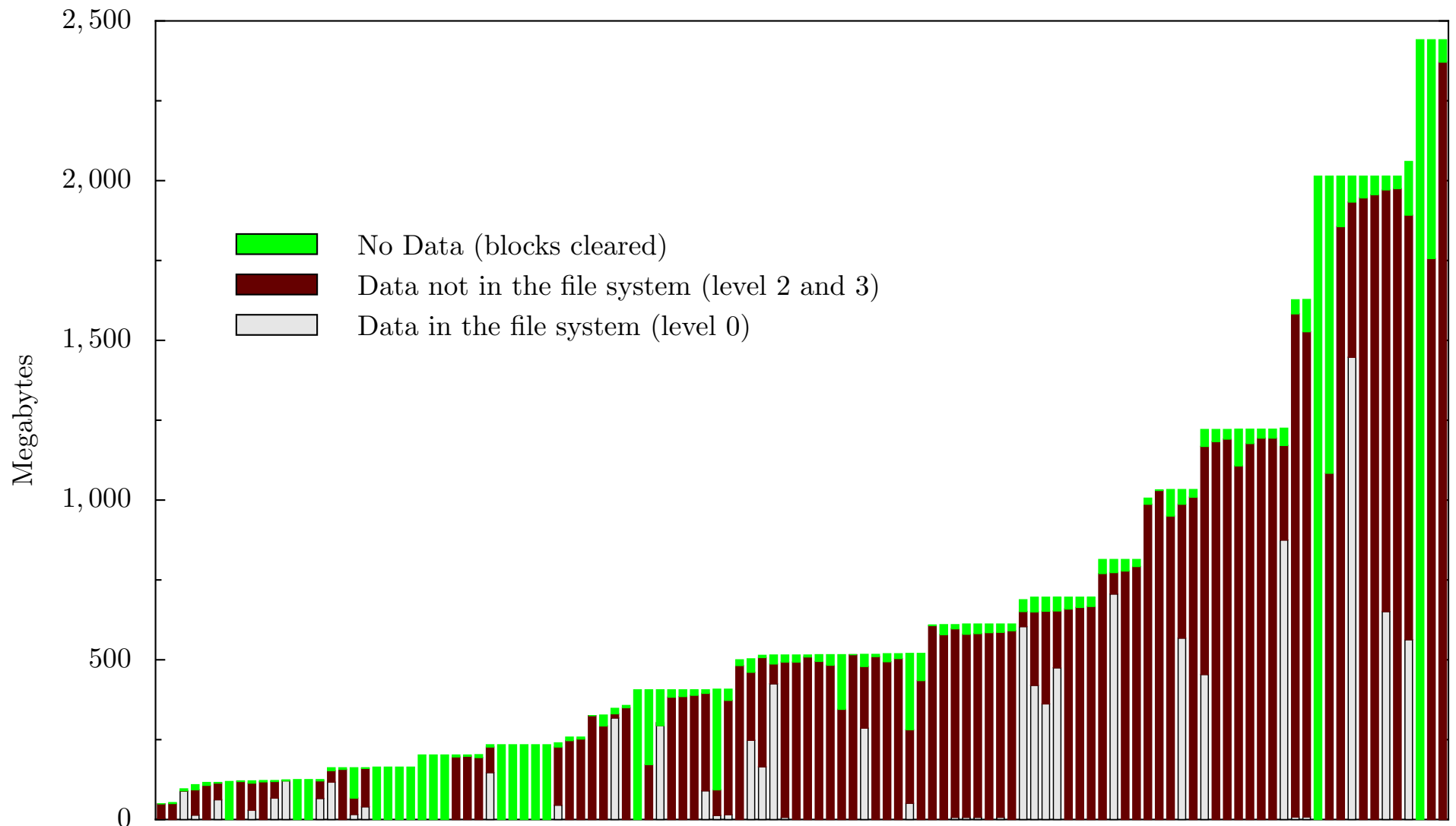
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:

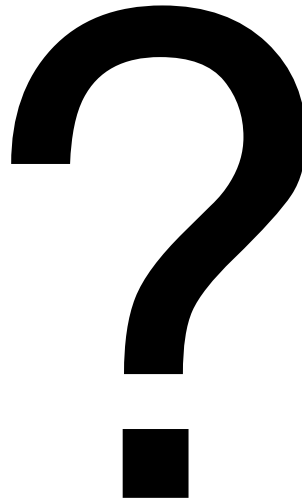


The 236 drives are dominated by failed sanitization attempts.



But training failures are also important.

But what *really* happened?



To answer this question, I needed to contact the original drive owners.

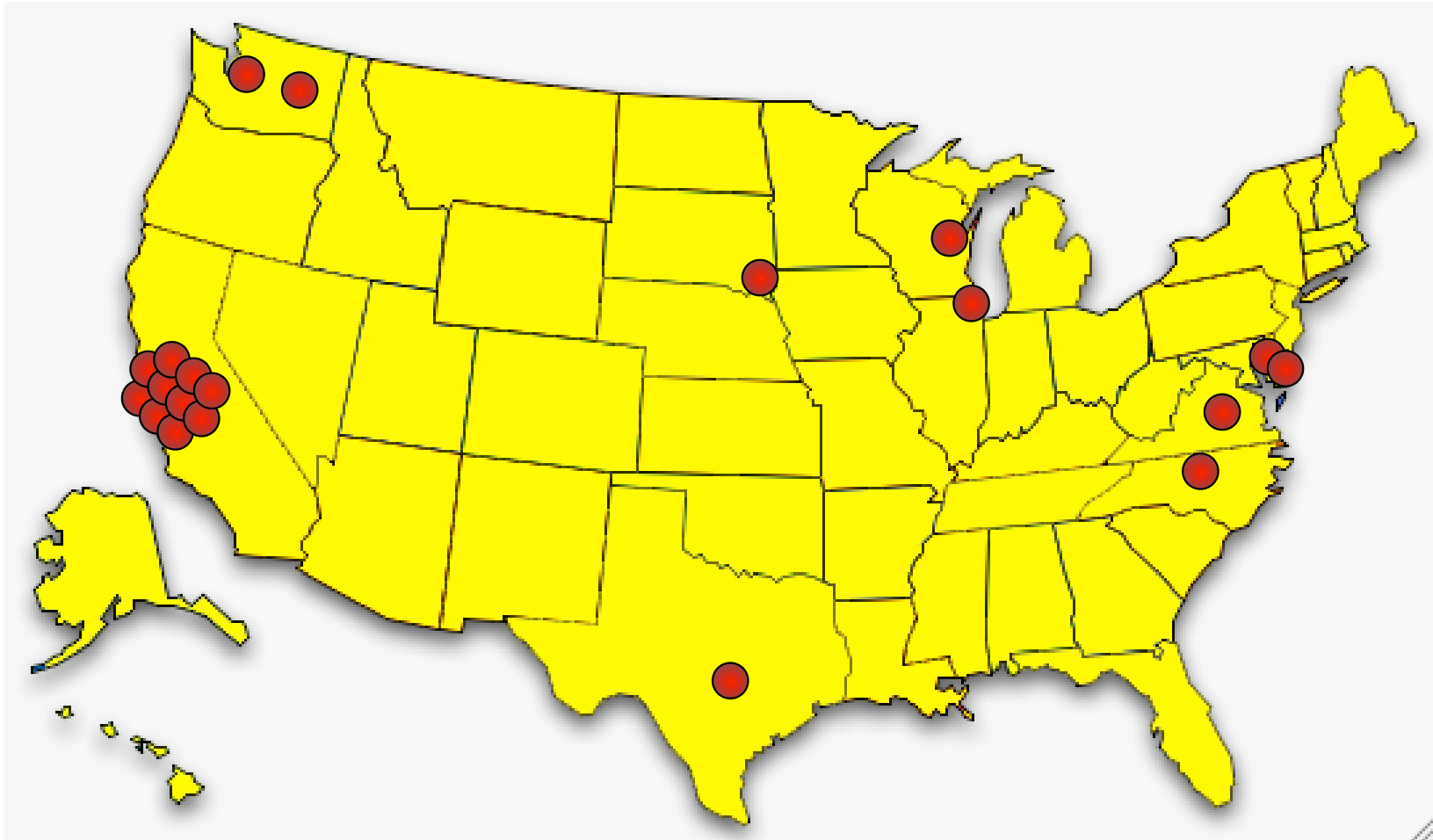
The Remembrance of Data Passed Traceback Study.

1. Find data on hard drive
2. Determine the owner
3. Get contact information for organization
4. Find the right person *inside* the organization
5. Set up interviews
6. Follow guidelines for human subjects work

```
06/19/1999 /:dir216/Four H Resume.doc
03/31/1999 /:dir216/U.M. Markets & Society.doc
08/27/1999 /:dir270/Resume-Deb.doc
03/31/1999 /:dir270/Deb-Marymount Letter.doc
03/31/1999 /:dir270/Links App. Ltr..doc
08/27/1999 /:dir270/Resume=Marymount U..doc
03/31/1999 /:dir270/NCR App. Ltr..doc
03/31/1999 /:dir270/Admissions counselor, NCR.doc
08/27/1999 /:dir270/Resume, Deb.doc
03/31/1999 /:dir270/UMUC App. Ltr..doc
03/31/1999 /:dir270/Ed. Coordinator Ltr..doc
03/31/1999 /:dir270/American College ...doc
04/01/1999 /:dir270/Am. U. Admin. Dir..doc
04/05/1999 /:dir270/IR Unknown Lab.doc
04/06/1999 /:dir270/Admit Slip for Modernism.doc
04/07/1999 /:dir270/Your Honor.doc
```

This was a lot harder than I thought it would be.

Ultimately, I contacted 20 organizations between April 2003 and April 2005.



The leading cause of compromised privacy was betrayed trust.

Trust Failure: 5 cases

- ✓ Home computer; woman's son took to "PC Recycle"
- ✓ Community college; no procedures in place
- ✓ Church in South Dakota; administrator "kind of crazy"
- ✓ Auto dealership; consultant sold drives he "upgraded"
- ✓ Home computer, financial records; same consultant

This specific failure wasn't considered in [GS 03]; it was the most common failure.

Poor training or supervision was the second leading cause.

Trust Failure: 5 cases

Lack of Training: 3 cases

- ✓ California electronic manufacturer
- ✓ Supermarket credit-card processing terminal
- ✓ ATM machine from a Chicago bank

Alignment between the interface and the underlying representation would overcome this problem.

In two cases, the data custodians simply didn't care.

Trust Failure: 5 cases

Lack of Training: 3 cases

Lack of Concern: 2 cases

- ✓ Bankrupt Internet software developer
- ✓ Layoffs at a computer magazine

Regulation on resellers might have prevented these cases.

In seven cases, no cause could be determined.

Trust Failure: 5 cases

Lack of Training: 3 cases

Lack of Concern: 2 cases

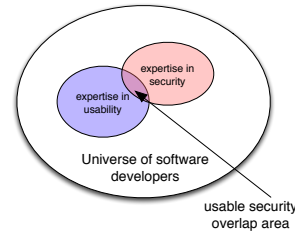
Unknown Reason: 7 cases

- ✗ Bankrupt biotech startup
- ✗ Another major electronics manufacturer
- ✗ Primary school principal's office
- ✗ Mail order pharmacy
- ✗ Major telecommunications provider
- ✗ Minnesota food company
- ✗ State Corporation Commission

Regulation might have helped here, too.

The Drives study establishes that data sanitization is a real problem.

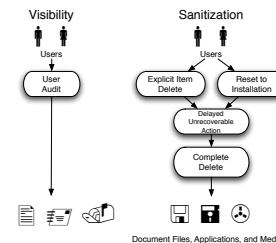
1. Introduction to patterns. ✓



2. The Drives Study ✓



3. Patterns for sanitization.

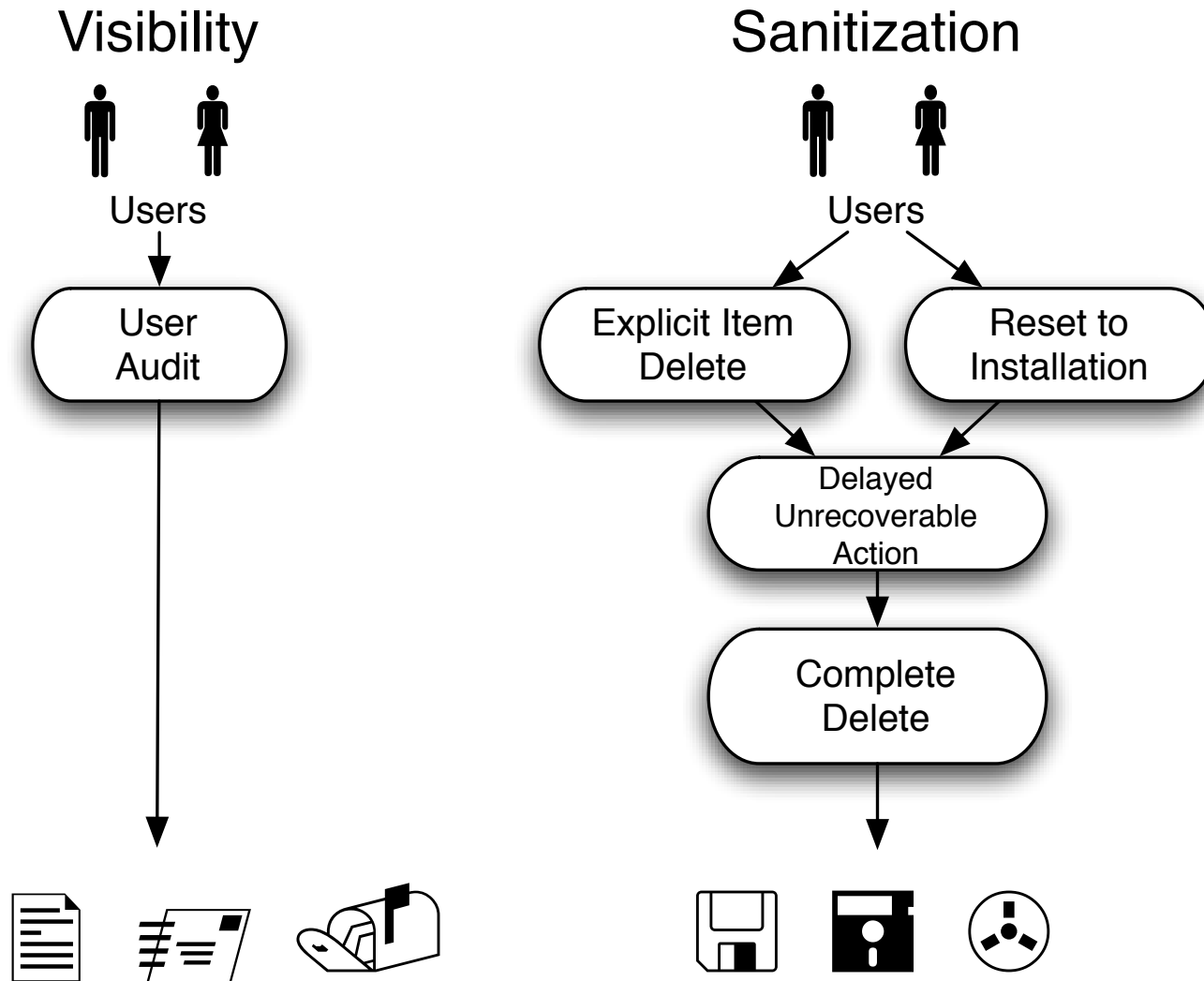


4. Applying these patterns.



But how can we solve it in a systematic fashion?

I have identified five distinct patterns for addressing the sanitization problem.



Document Files, Applications, and Media

Complete Delete: assure that deleting the *visible* representation deletes the *hidden* data as well.

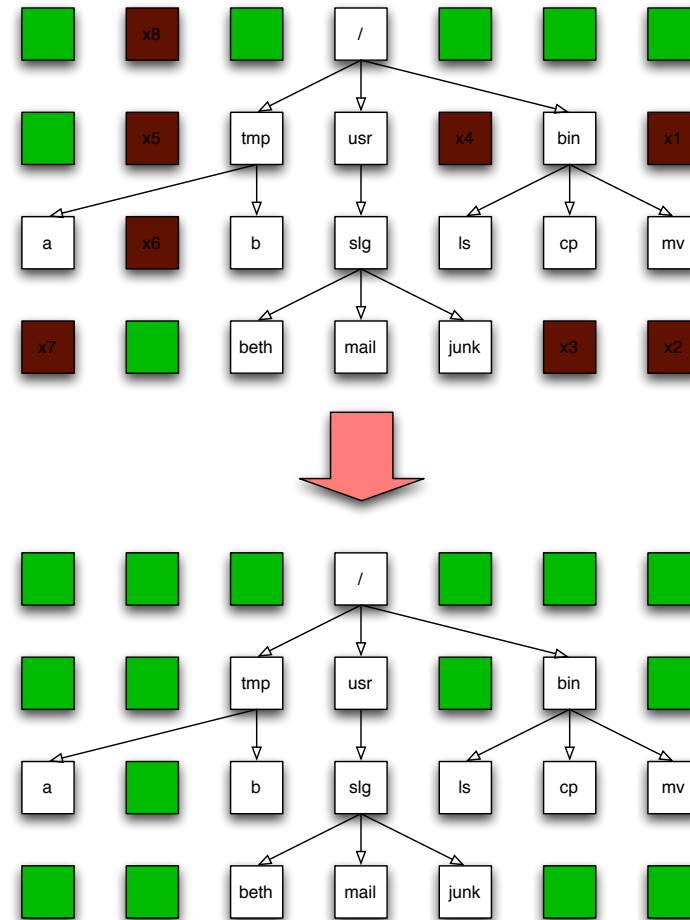
Sanitization



Complete
Delete



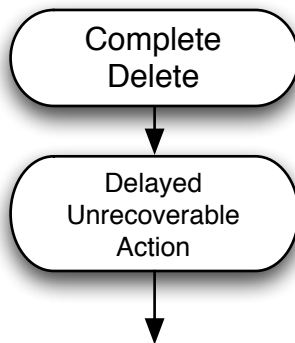
Document Files, Applications, and Media



Naming this pattern lets us discuss its absence in modern operating systems.

Delayed Unrecoverable Action: give the users a chance to change their minds.

Sanitization

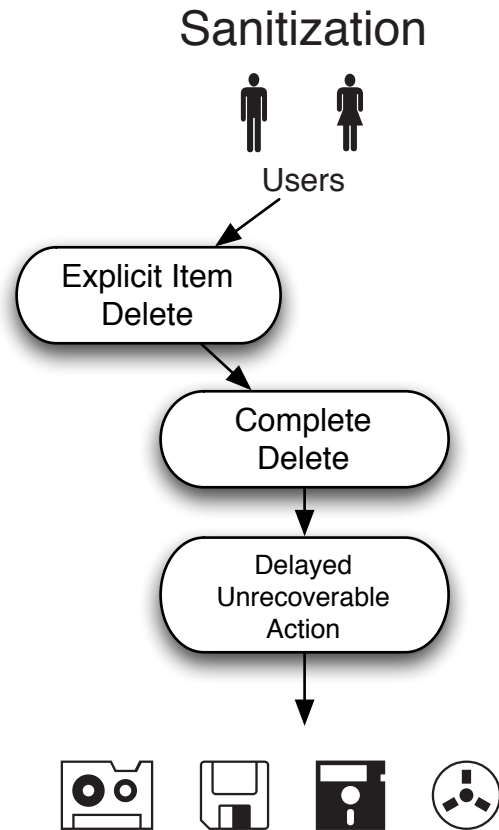


Document Files, Applications, and Media



[Norman 83] and [Cooper 99] both suggest this functionality, but they do not give it a name.

Two ways to delete information. #1: *Explicit Item Delete*



Document Files, Applications, and Media

```
C:\WINDOWS\system32\cmd.exe

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

Directory of C:\tmp

10/15/2004  09:20 PM    <DIR>          .
10/15/2004  09:20 PM    <DIR>          ..
10/03/2004  11:34 AM                27,262,976 big_secret.txt
               1 File(s)                27,262,976 bytes
               2 Dir(s)         4,202,078,208 bytes free

C:\tmp>del big_secret.txt

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

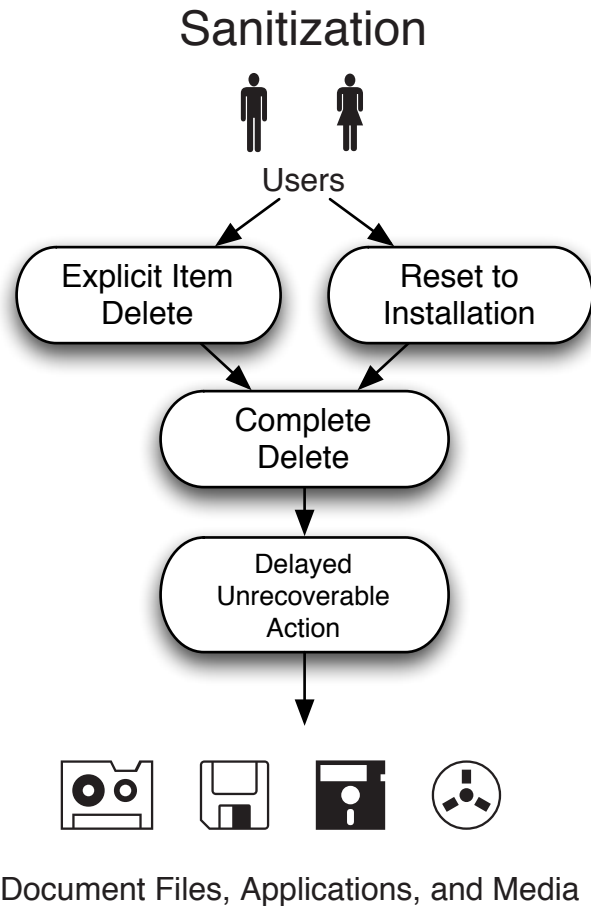
Directory of C:\tmp

10/15/2004  09:22 PM    <DIR>          .
10/15/2004  09:22 PM    <DIR>          ..
               0 File(s)                   0 bytes
               2 Dir(s)         4,229,296,128 bytes free

C:\tmp>_
```

“Provide a means for deleting information where the information is displayed.”

Reset to Installation: Get rid of everything



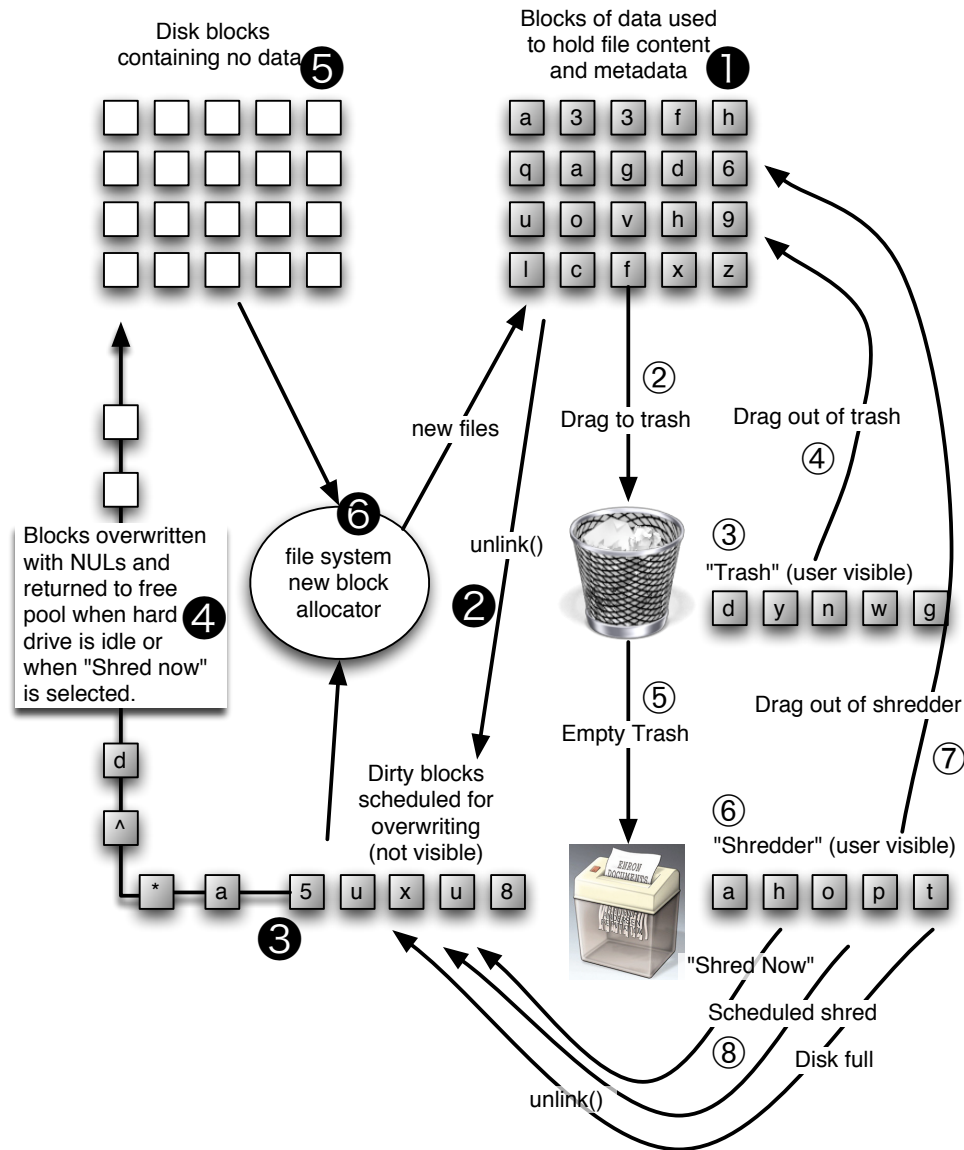
```
C:\>C:\WINDOWS\system32\cmd.exe - format c:

C:\>format c:
The type of the file system is NTFS.

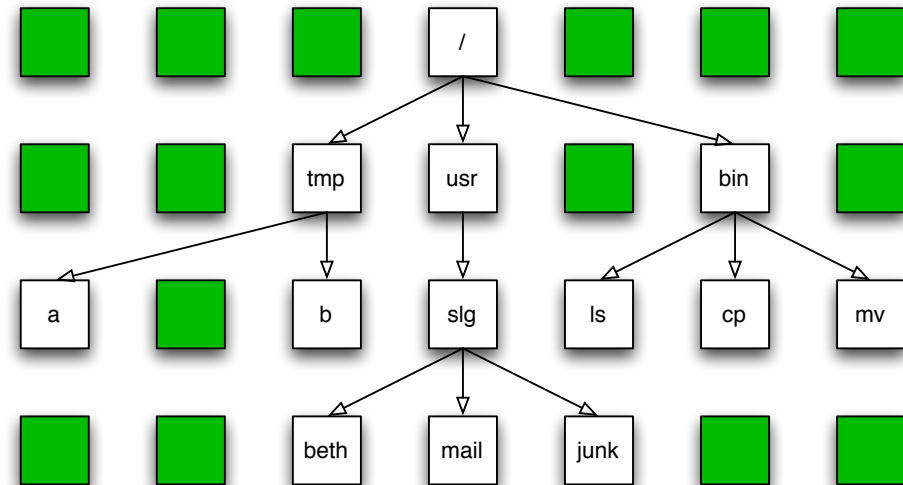
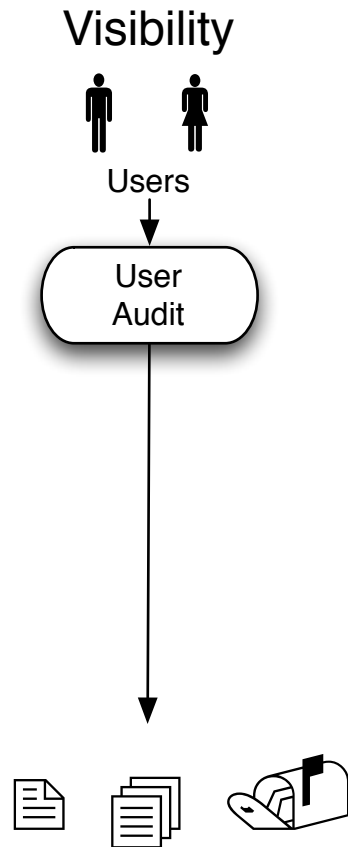
WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

**Reset/reinstall functionality is common (Windows; PalmOS; etc.).
This pattern framework clarifies *Reset's* security property.**

The shredder should be integrated into the file system.

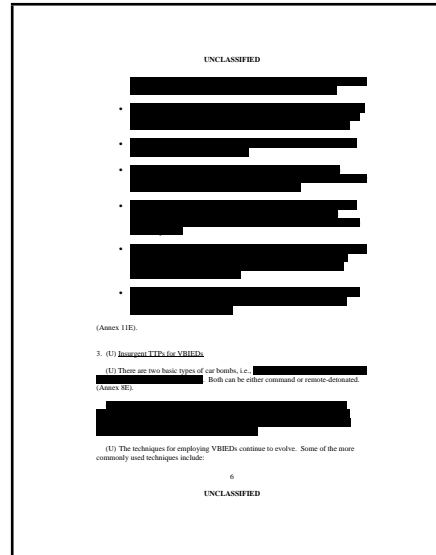


User Audit: If the information is present, make it visible.



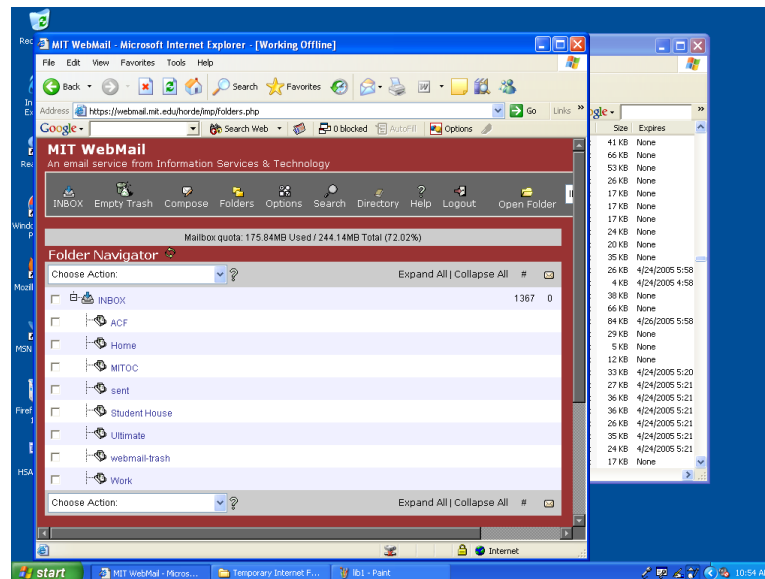
With files, this happens automatically
when the *Complete Delete* pattern is implemented.

The power of these patterns is that they apply equally well to other sanitization problems.



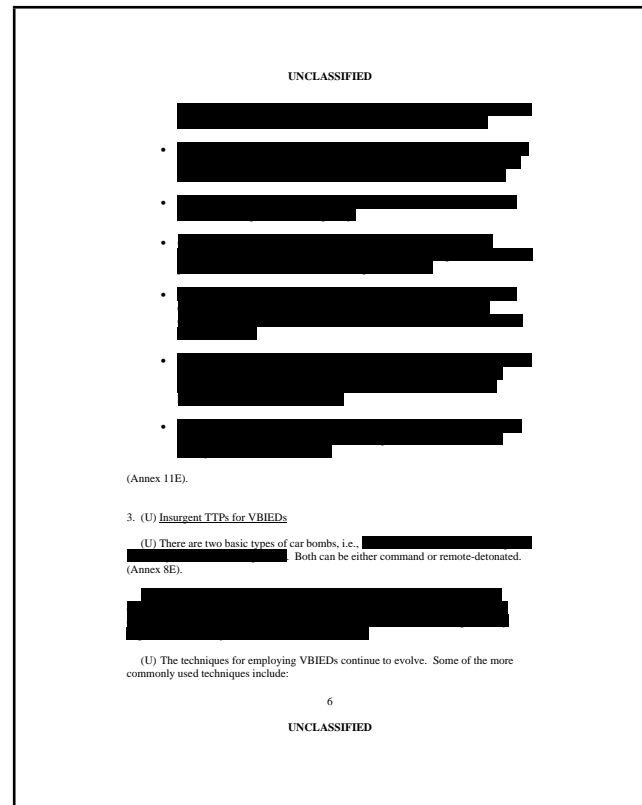
- Document Files

- Web Browsers



Information is left in document files.

- The *New York Times* published a **PDF file** containing the names of Iranians who helped with the 1953 coup. [Young 00]
- US DoJ published a **PDF file** “diversity report” containing embarrassing redacted information. [Poulsen 03]
- SCO gave a **Microsoft Word file** to journalists that revealed its Linux legal strategy. [Shankland 04]
- Multinational Forces-Iraq (MNF-I) report on the death of Nicola Calipari, March 4, 2005



Information was unintentionally hidden in the MNF-I report.

UNCLASSIFIED

- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

(Annex 11E).

3. (U) Insurgent TTPs for VBIEDs

(U) There are two basic types of car bombs, i.e., [REDACTED]. Both can be either command or remote-detonated. (Annex 8E).

[REDACTED]

(U) The techniques for employing VBIEDs continue to evolve. Some of the more commonly used techniques include:

6

UNCLASSIFIED

UNCLASSIFIED

easy to emplace by staging equipment in vehicles or near overpasses, and, in a matter of minutes, having the IED armed and in the desired location.

- (S//NF) Explosives wrapped in a brown paper bag or a plastic trash bag. This is a particularly easy method of concealment, easy to emplace, and has been used effectively against Coalition Forces and civilians along Route Irish.
- (S//NF) Explosives set on a timer. This technique is new to the Route Irish area, but is being seen more frequently.
- (S//NF) Use of the median. The 50 meter wide median of Route Irish provides a large area for emplacing IEDs. These can be dug in, hidden, and/or placed in an animal carcass or other deceptive container.
- (S//NF) Surface laid explosives. The enemy will drop a bag containing the explosive onto the highway and exit the area on an off-ramp with the detonation occurring seconds or minutes later depending on the desired time for the explosion.
- (S//NF) Explosives on opposite sides of the median. Devices have been found along both sides of the median that were apparently designed to work in tandem, to counter Coalition Force tactics to avoid the right side of the highway while traveling Route Irish.
- (S//NF) Explosives hidden under the asphalt. Insurgents pretend to do work on the pavement, plant the explosives, and repair the surface. These are usually remote-detonated devices.

(Annex 11E).

3. (U) Insurgent TTPs for VBIEDs

(U) There are two basic types of car bombs, i.e., suicide (where the car is moving) and stationary (where the car is parked). Both can be either command or remote-detonated. (Annex 8E).

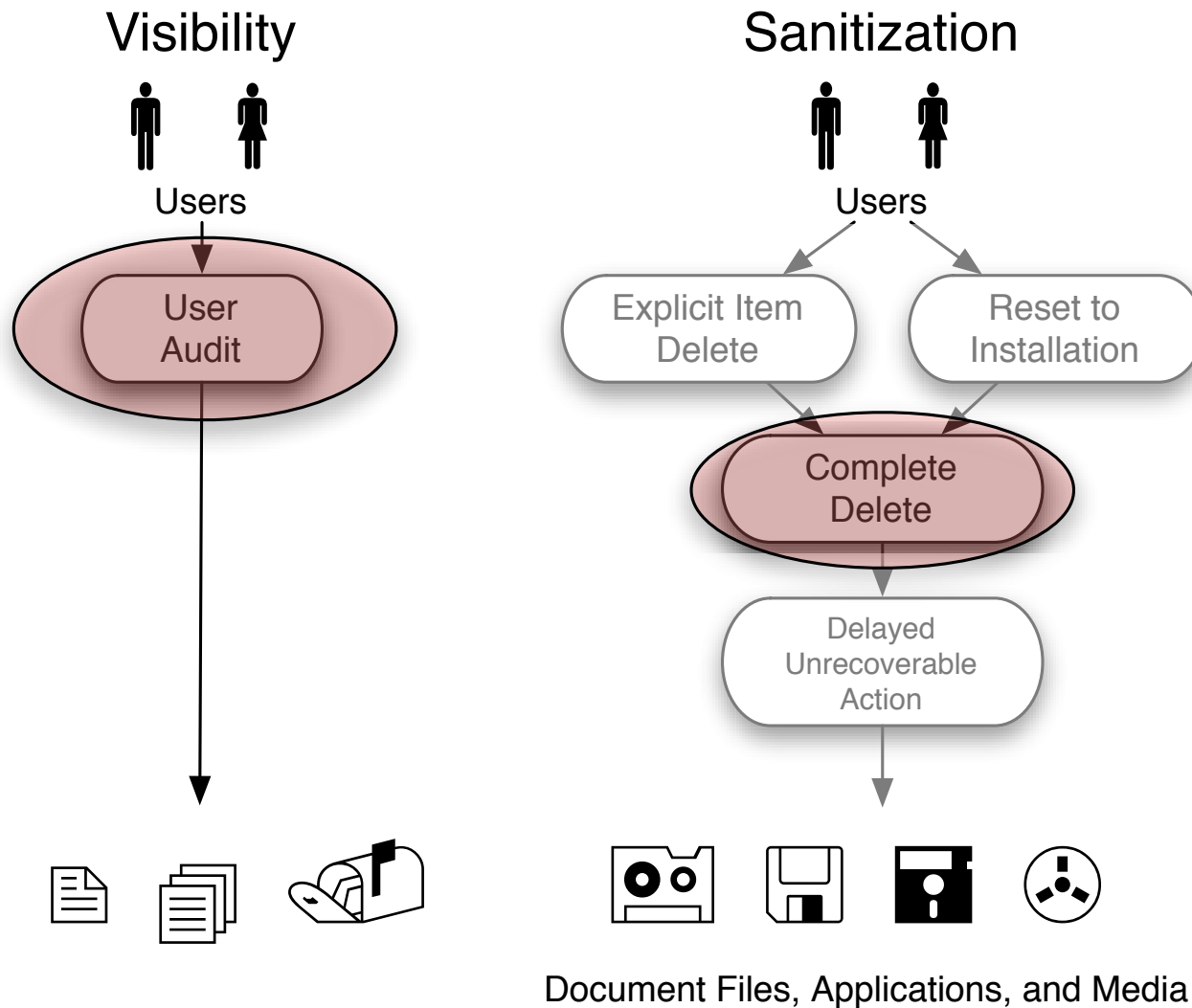
(S//NF) The enemy is very skillful at inconspicuously packing large amounts of explosives into a vehicle. The most commonly used detonation materials are plastic explosives and 155mm artillery shells. When moving, these VBIEDs are practically impossible to identify until it is too late. (Annex 8E).

(U) The techniques for employing VBIEDs continue to evolve. Some of the more commonly used techniques include:

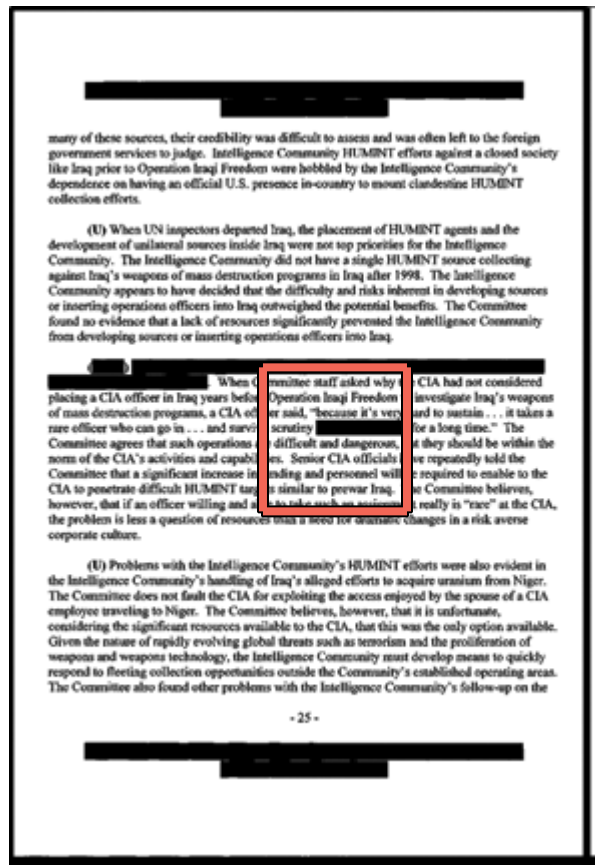
6

UNCLASSIFIED

The information leaked because two patterns were not implemented.

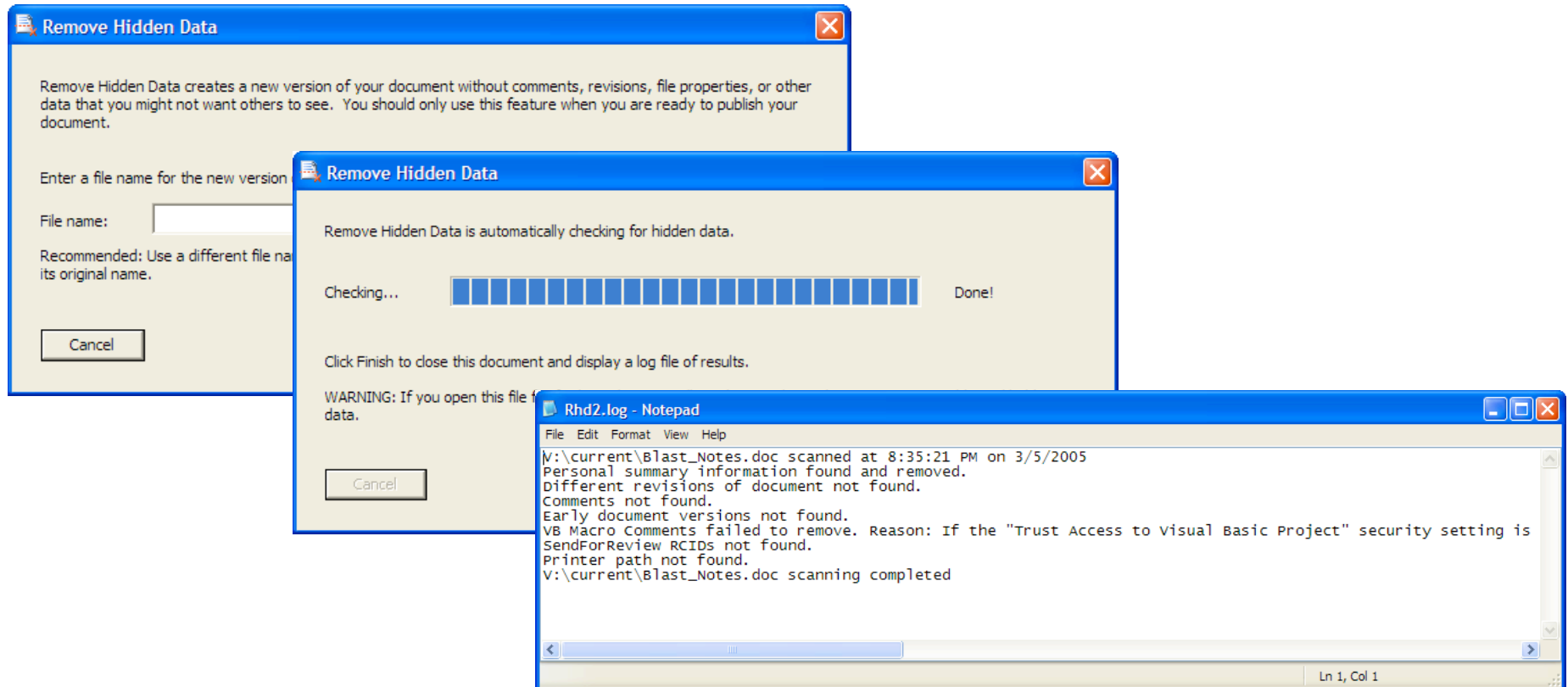


The Senate Foreign Intelligence Committee accomplished this goal by *scanning* the redacted report on pre-war Iraq intelligence to create the PDF that it distributed.



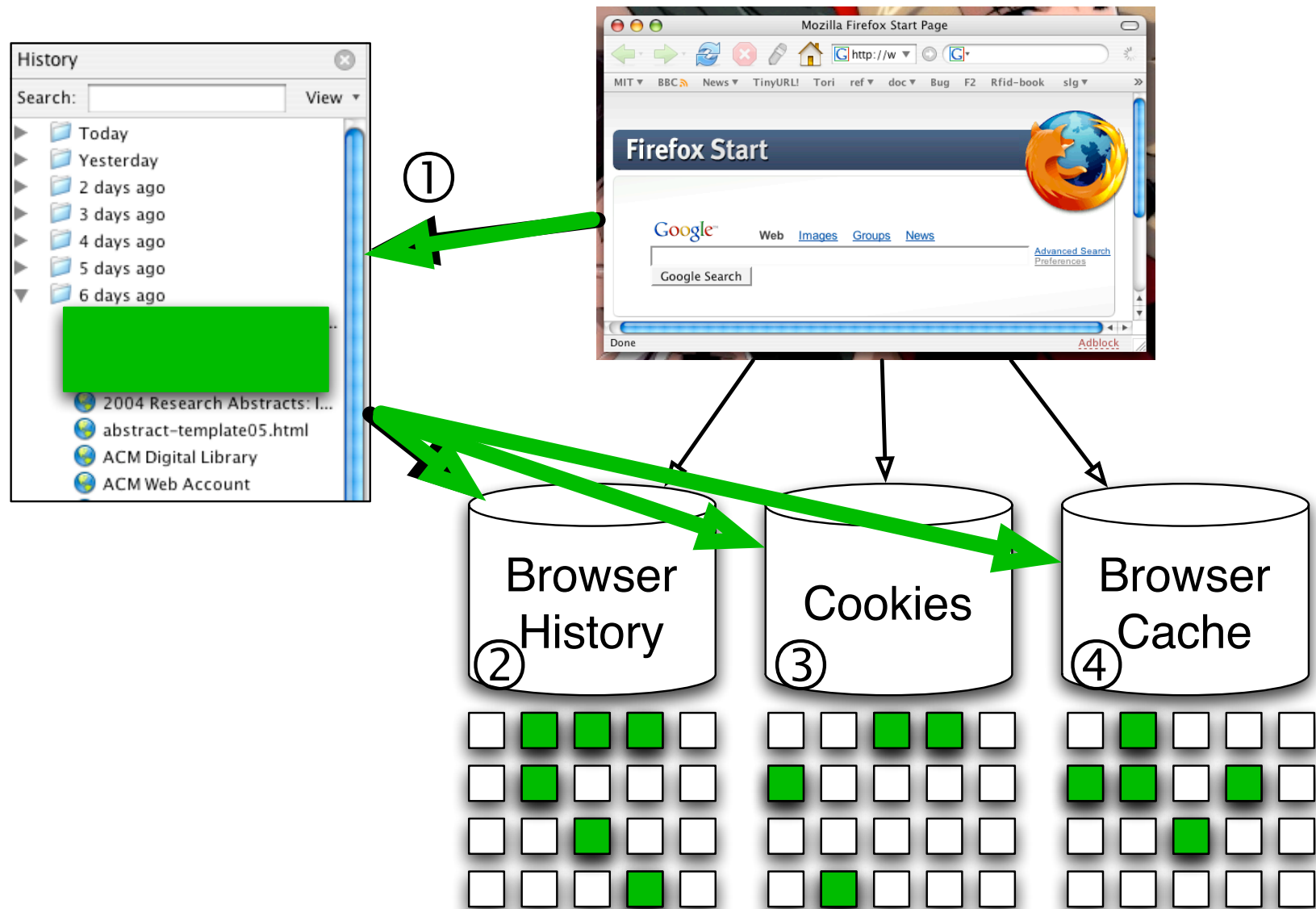
nittee staff asked w
eration Iraqi Freed
said, "because it's
rutiny [redacted]
difficult and danger
Senior CIA offic
ling and personnel
similar to prewar Ir

Microsoft has tried to solve this problem with “Remove Hidden Data” tool.



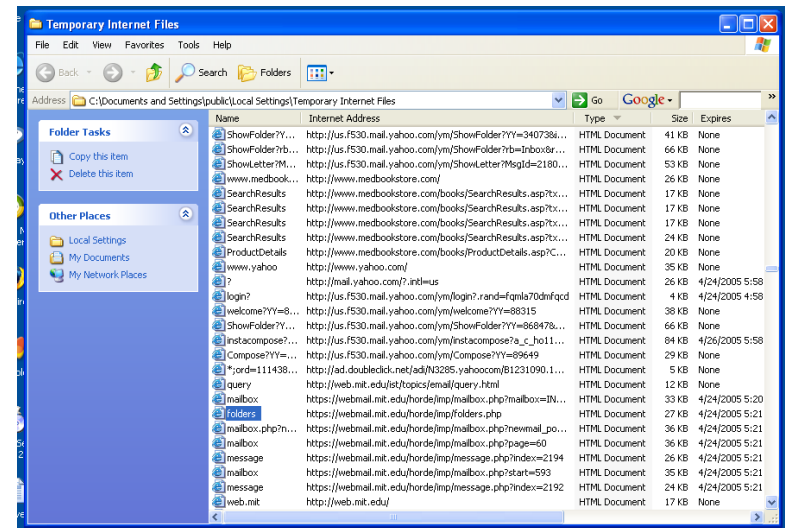
RHD doesn't integrate into the flow of document preparation. The patterns-based analysis predicts that RHD will fail in many cases.

Information is left behind in web browsers.



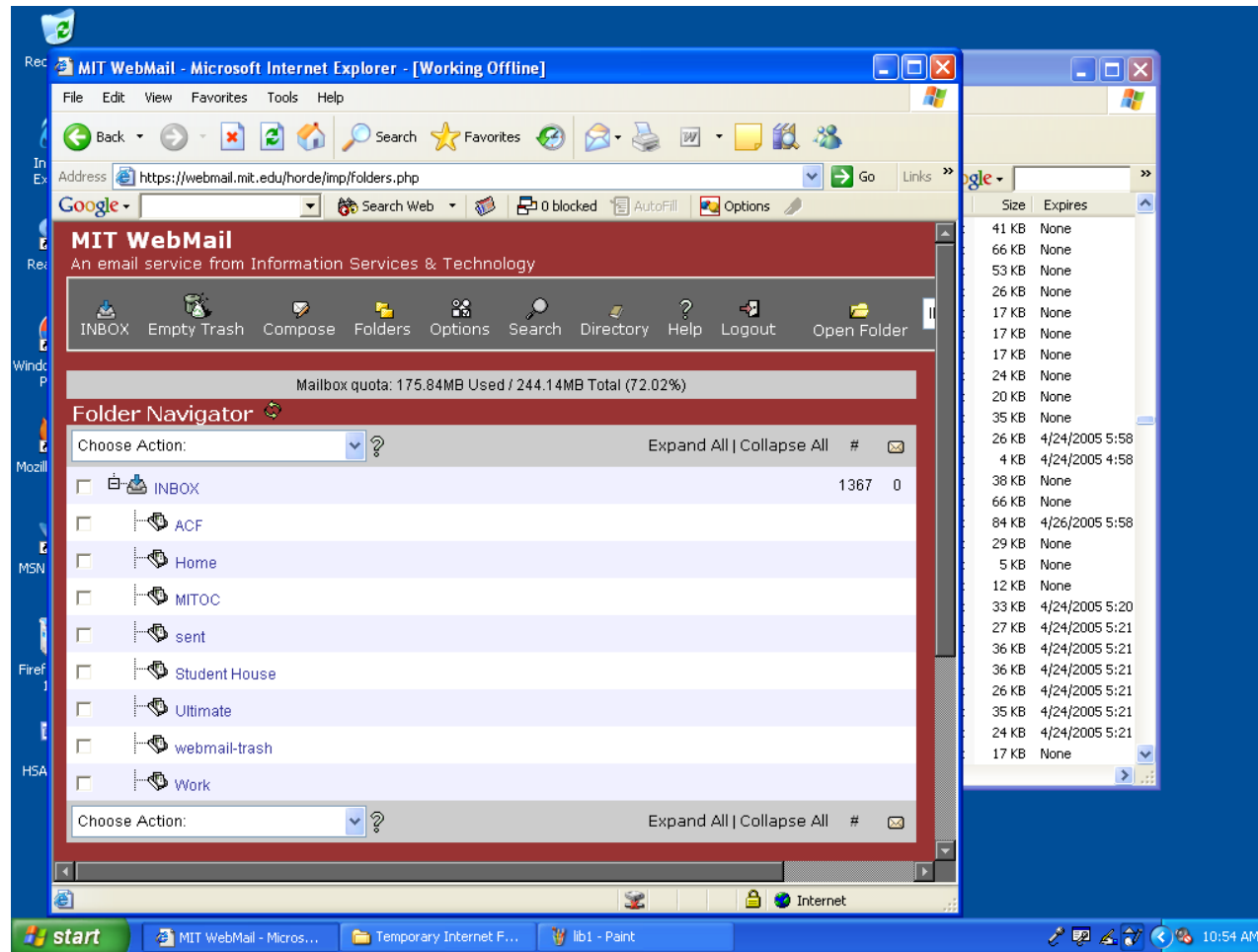
Two key problems: ① Deleted files; ② The cache

In fact, a lot of information is left behind in web browsers.



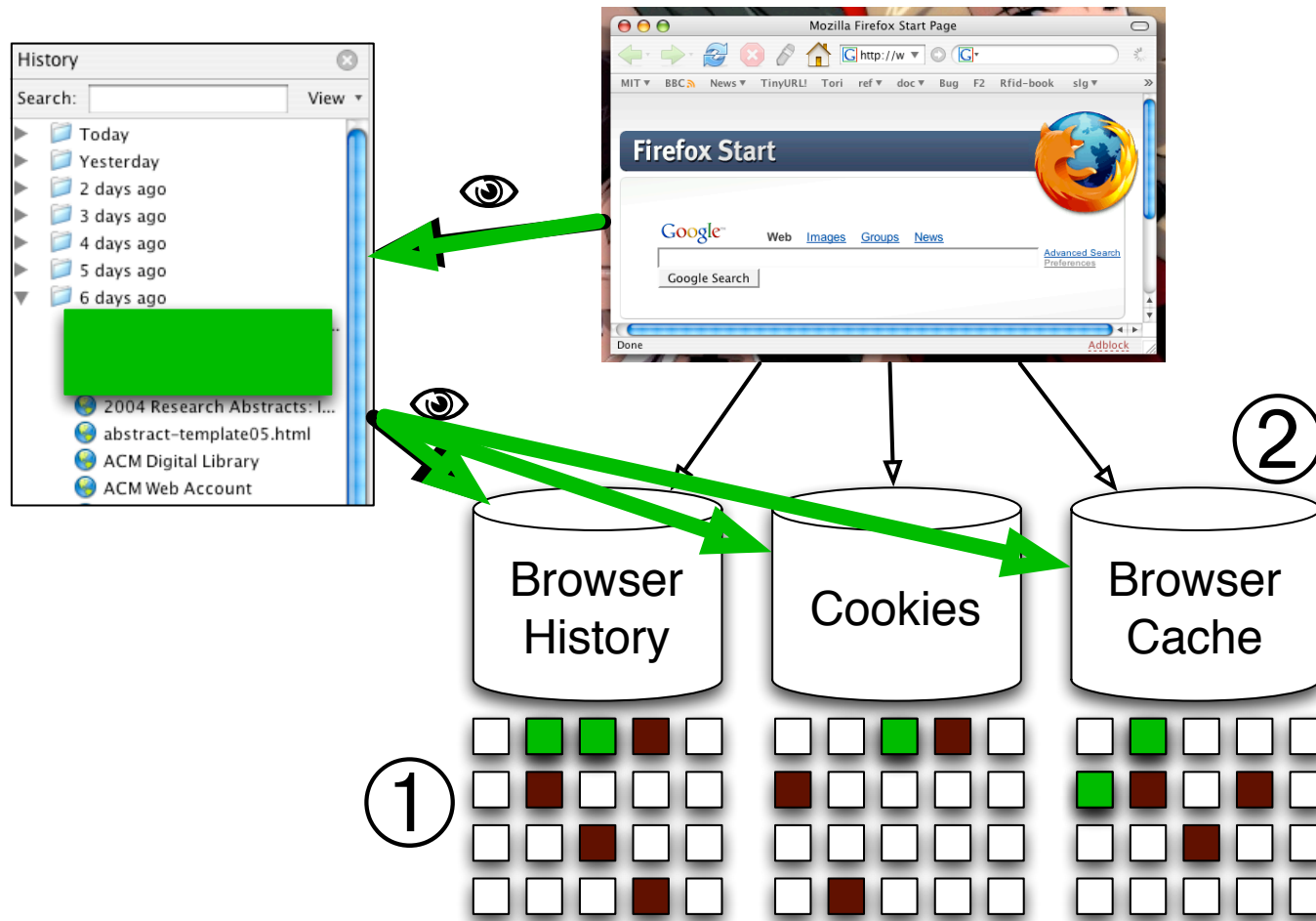
MIT Humanities Library, April 25, 2005

4 out of 4 computers inspected had significant quantities of personal email in their browser caches.



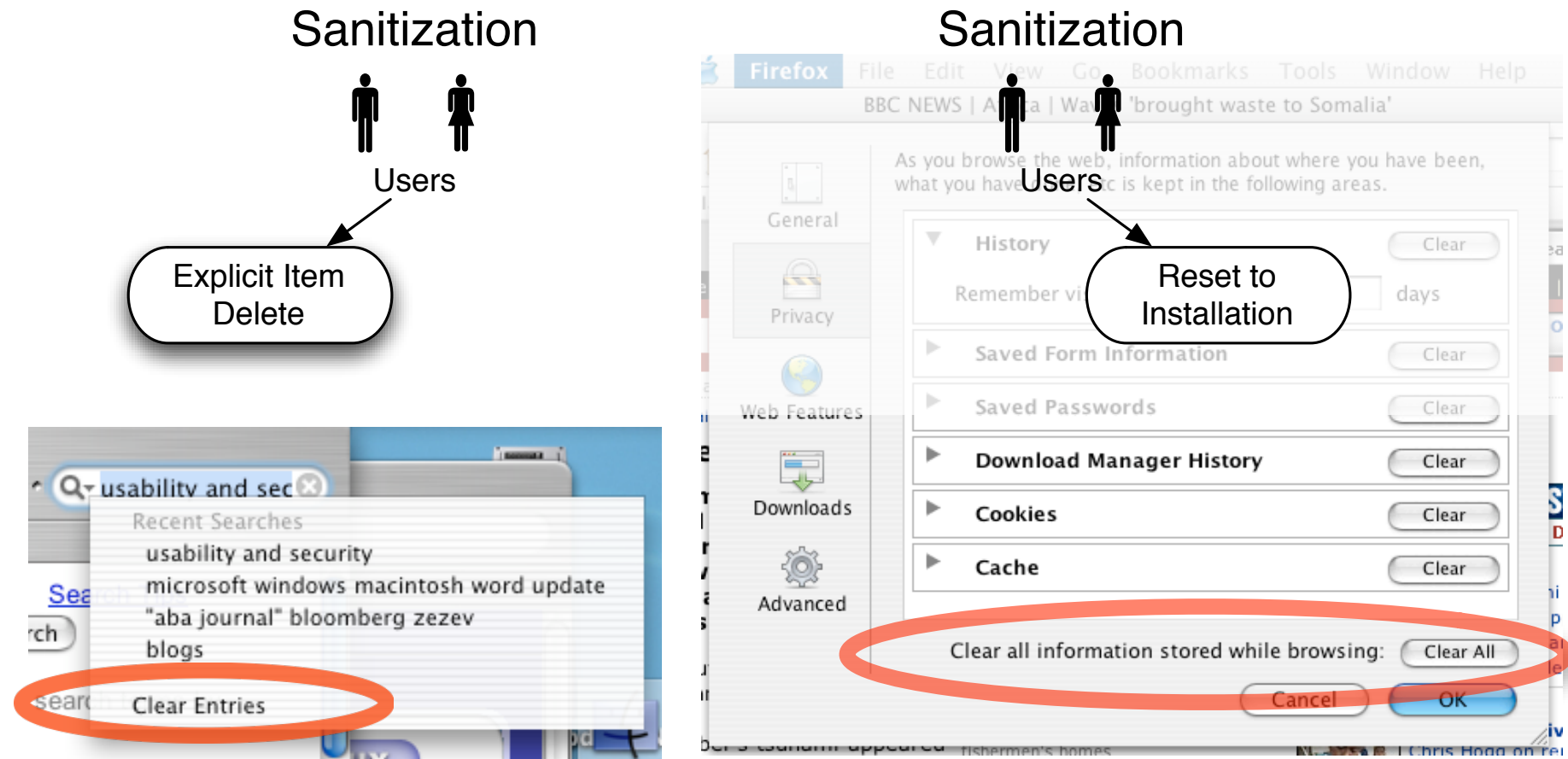
The American Library Association recommends software that automatically purges caches on a *daily* basis.[ALA 05] (It would be better to purge after each use.)

**Applying the patterns,
an obvious solution is to unify the history and cache:**



**The patterns make it easy to explain this concept to the
browser developers and users.**

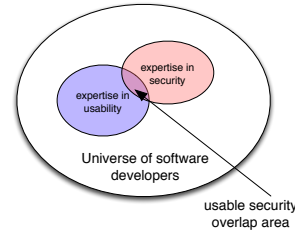
The patterns also suggest opportunities for further promoting HCI-SEC within the browser.



Without *Complete Delete* the data can still be recovered. This demonstrates the need for the complete pattern set.

In Summary

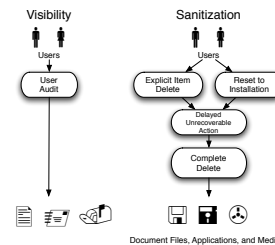
1. Introduction to patterns. ✓



2. The Drives Study ✓



3. Patterns for sanitization. ✓



4. Applying these patterns. ✓



Questions?