

THE DATA SERIES, FROM TALLY TO TOKEN

Data on Every Desk

Files, Spreadsheets and the PC Era

The personal computer made the record a file a person could create, copy and rename, and made every copy editable, ownerless and indistinguishable from the original. This book follows one hotel group's rate workbook from a Lotus file on a floppy in 1987 to a workplace assistant in 2026, and gives a CTO what to do before an assistant reads the desk: name the authoritative copy, count the copies, narrow who can open them, retire and delete for real, and keep the Ledger as evidence of which copy an answer used.

Juan Cardena

Enterprise AI & Agentic Systems Architect · blog.jcardena.com

CONTENTS

BEFORE YOU START

Reader's Brief: One Sheet, Six Questions

PROLOGUE

One Sheet, Since 1987

PART I · THE FILE

01 The Command Line and the File

02 What a File Knows and What It Forgets

PART II · THE DESK TOOLS

03 VisiCalc, Lotus 1-2-3, Excel: the Grid as a System

04 When the Spreadsheet Is the System

05 dBase and Access: the Desktop Database

PART III · THE COPY SPREADS

06 Shared Drives, Email and the Copy as Default

07 CSV, the Format That Never Died

08 Client/Server and the Pipe to the Desk

PART IV · CLOUD AND POCKET

09 From Shared Drive to Cloud Drive

10 The Copy That Left the Building

PART V · SHADOW AND THE NEW READER

11 Deleting a File for Real

12 Shadow Data: Finding Every Copy

13 The Assistant Reads Your Files

PART VI · THE WORK

14 Authority and Inventory

15 Permissions, Retirement and Proof

CLOSING

Epilogue: The Copy Outlives the Desk

APPENDICES

A The Ledger for Files: Template, Inventory and Readiness Check

B Timeline

C Failure Library

D Glossary

E Sources

BEFORE YOU START

Reader's Brief: One Sheet, Six Questions

This book follows one hotel group's rate workbook from a Lotus file on a floppy disk in 1987 to a workplace assistant's answer in 2026. It asks the same six questions of every copy the workbook left behind, and it gives you three tools: the Data Lineage Ledger, the authority states and a readiness check for assistants.

Three companies, all illustrative

- **The hotel and retail group:** about 60 hotels and 400 stores in 2026. In 1987 its predecessor ran 12 hotels. It owns the thread: the rate sheet. Processing for its EU hotels falls under European data protection law.
- **The insurer, founded 1898:** a reserving workbook, a claims tracker built in a desktop database in 1996, and loss adjusters' tablets.
- **The health system:** a clinic waiting list kept in a spreadsheet, and medical record numbers that lose their leading zeros.

Scope: every file on the desk

Spreadsheets are the thread, because pricing, reserving and reporting logic often lives in them, outside any system. But every desk file counts: documents, PDFs, desktop databases and the exports that feed them. Where this book says "file", read any of them.

Six questions asked of a file

- **Where is it?** Path, platform and every copy.
- **What does it mean?** Columns, codes, formulas and units.
- **Who can change it and who can open it?** Rights, groups and links.
- **What does it cost?** To store, find, keep and rework.
- **Who reads it now?** People, macros, uploads and assistants.
- **What proves it is the right copy?** The authority record (Chapter Two).

Each move of a copy, or change in who can reach it or whether it still counts, is a **hop**, numbered 0 to 10 in time order; a lettered hop, such as 1b or 5b, comes after the hop with that number and before the next. Each hop is one Ledger row with ten columns, and its loss codes name what the move lost: MEANING (a code or definition did not travel with the value), CONTEXT (notes or surrounding knowledge stayed behind), PRECISION (type or

rounding changed the value), TIME (century, time zone, or event versus processing time changed), IDENTITY (the key was renumbered, merged or split), OWNERSHIP (a copy has no accountable owner), RIGHTS (purpose, consent or legal basis did not travel) and COMPLETENESS (rows or fields were dropped). One row shows how a single move changed the answer to the third question:

THE HOP · 5 · 2016 · WINDOWS FILE SERVER (GROUPS REMAPPED FROM NETWARE TRUSTEES IN 2006) TO SHAREPOINT ONLINE LIBRARY REVENUE

Record	The rate sheet (hotel and retail group), hop 5, 2016
From	Windows file server (groups remapped from NetWare trustees in 2006)
To	SharePoint Online library Revenue
Mechanism and schedule	A third-party migration tool moves the share to the cloud; to finish on time the team maps unresolved and nested groups to one custom group
Format change; what was lost or assumed	RIGHTS Files start a version history in the library; permissions inherited from the library; no item keeps narrower permissions. **The custom group All staff (internal), then the hotel and head office staff (about 4,000 people), gets read access to the Revenue library; the corporate tab becomes readable by all of them**
Accountable owner	Head of revenue (site owner)
Who can change it	The revenue team (edit); site collection administrators
Classification; retention or erasure duty	Personal data and commercial terms; no retention policy set on the library
Read by (people, apps, models, agents)	All staff (internal) (read): about 4,000 staff in 2016, few of whom need the CORP tab
Cost per month; basis	One-off: the migration project (not monthly); per month: a share of licences (illustrative)

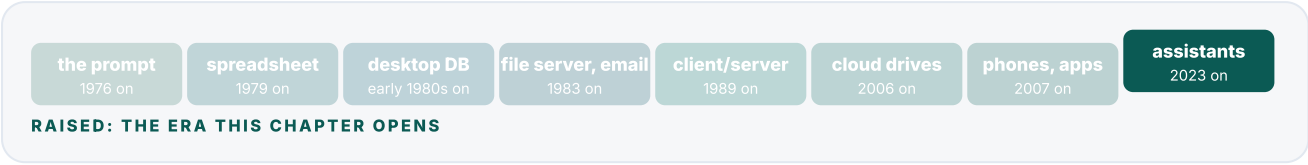
Reading paths by role

Role	Read first	The decision you leave with
CTO	Prologue, Chapters 1, 13, 14, 15	What must be true before an assistant reads the desk
Chief data officer	Chapters 2, 8, 12, 14, Appendix A	Which desk files are systems of record in all but name
Data or AI architect	Chapters 2, 3, 7, 8, 13	Which formats and rules a file must meet to enter a system
CISO	Chapters 9, 10, 12, 13, 15	Which groups and links to narrow, and in what order
Records manager	Chapters 2, 6, 11, 12, Appendix A	How retirement and deletion reach every copy

Monday rule: pick one record your business prices or reports from, and try to count its copies and name who can open each one.

PROLOGUE

One Sheet, Since 1987



On Tuesday 10 March 2026 a retail B2B account manager at the hotel and retail group asked the workplace assistant a simple question: what is the best rate at the harbor hotel for a client's festival weekend? The answer came back in seconds, with a citation. It did two kinds of harm: it quoted a rate that did not apply, and it put two clients' names and contacts in front of a user who then sent them outside the group. The assistant bypassed no permission to give it. The decision it changes: **before an assistant reads your files, every important record has a named authoritative copy, and every other copy is either retired in a way a machine can read or out of reach**.

Tuesday: the question and the answer

The account manager was one of 300 people in the group's assistant pilot. The harbor hotel is one of the group's EU hotels. The assistant answered 189.00 euros, and it named two corporate clients and their contacts as holders of that rate. It cited its source: `Rates master.xlsx` in the Revenue library.

The answer looked complete: a number, a reason and a document behind it. The account manager pasted it into an email to the client and sent it.

THE HOP · 10 · 2026 · ASSISTANT ANSWER TO EMAIL TO A CLIENT (A RECIPIENT-HELD COPY OUTSIDE THE GROUP, AND A SENT ITEM INSIDE)

Record	The rate sheet (hotel and retail group), hop 10, 2026
From	Assistant answer
To	Email to a client (a recipient-held copy outside the group, and a sent item inside)
Mechanism and schedule	The account manager pastes the answer into an email on Tuesday 10 March 2026
Format change; what was lost or assumed	TIME RIGHTS Email text with a rate and two clients' names and contacts. **A corporate rate nearly four years old that does not apply on festival dates (189.00 euros; the applicable rate is the public 236.00) and two clients' names and contacts leave the group**
Accountable owner	Retail division sales
Who can change it	The account manager
Classification; retention or erasure duty	Personal data disclosed to a third party; commercial terms
Read by (people, apps, models, agents)	The client and anyone they forward it to
Cost per month; basis	One-off: lost margin on one quote and a breach assessment (not monthly) (illustrative)

That email is a hop in the Ledger: a copy held by the recipient. Two things travelled with it that should not have. The first was a corporate rate nearly four years old that does not apply on festival dates. The second was the names and contacts of two corporate clients, whose terms with the group are confidential.

Thursday: the reply and the trace

On Thursday 12 March 2026 the client replied. It quoted 189.00 euros and asked for the same terms as one of the named companies. The revenue manager began to trace the answer.

The revenue manager opened the cited workbook read-only and read its file properties. It was last saved in June 2022. The assistant's citation had said where the number came from; only the file properties said how old it was. The applicable rate was in the revenue management system: 236.00 euros, the public rate for those dates. The 189.00 euros was the

harbor hotel's corporate rate as last maintained by hand in June 2022. The query that once filled the CORP tab, the workbook's first sheet, died in 2010; its client contacts stayed as they were then, and its rates were typed by hand until that last save.

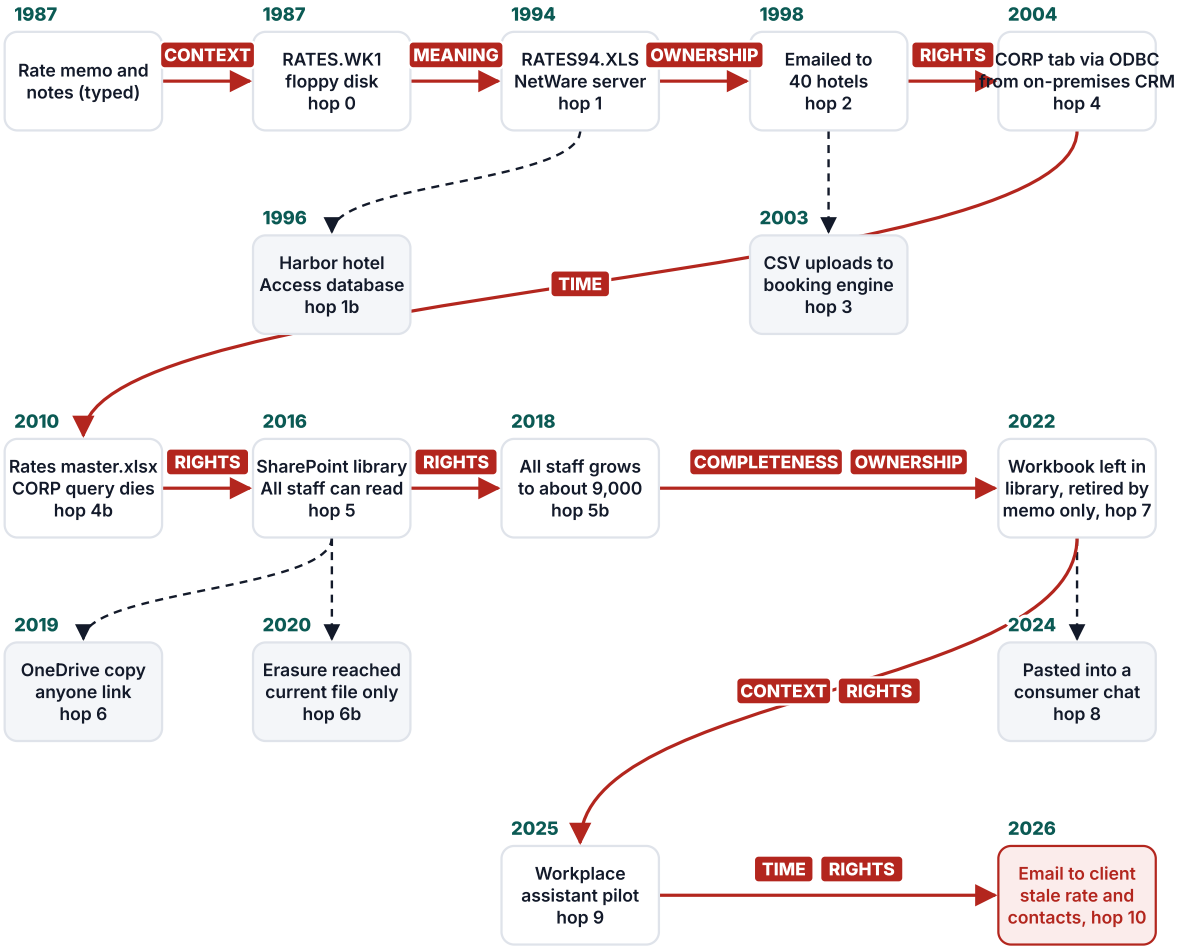
Corporate rates carry a festival blackout. That condition has never lived in the workbook or in the revenue system. It began in a typed memo in 1987, and today it is a clause in each corporate agreement held in the CRM, a rented service since 2010 and the home of the corporate agreements of record. Neither the revenue system nor the CRM is connected to the assistant.

Six names back to a floppy disk

Counting each path as a name, the walk back found six: five on the causal chain and one copy off it. `Revenue/Rates/Rates master.xlsx` had sat in a SharePoint Online library since a migration in 2016. Before that it was `Rates master.xlsx` on a Windows file server, saved in the new Excel format in 2010. Before that it was `S:\REVENUE\RATES94.XLS` on a NetWare server (a Windows server from 2006), rebuilt in Excel in 1994, with copies emailed to every hotel from 1998 under names such as `RATES98 final GM.xls`. At the bottom of the chain was `RATES.WK1`, a Lotus 1-2-3 worksheet that the reservations manager typed on an MS-DOS PC in 1987 and kept on a floppy disk. The festival blackout was already missing then: it stayed in the typed memo, and the sheet held only codes and numbers.

FIGURE P.1

One sheet, since 1987: sixteen boxes, the causal path in solid boxes, the side branches in pale ones



Red arrows and chips mark a hop where something was lost (loss codes, Chapter 2).

The rate sheet from a Lotus file on a floppy disk to an assistant's answer in an email: solid boxes are the hops that led to the 2026 answer, pale boxes are side branches (their loss codes are in their chapters), and each red chip is a Ledger loss code.

Author's diagram based on the book's running example; the hotel and retail group, its people and numbers are illustrative.

The sixth name belonged to a copy off the causal chain. In 2019 a general manager had saved `rates final GM v2.xlsx` to her OneDrive and sent an events agency a link that needs no sign-in and has no expiry. That copy is outside the library's permissions and retention configuration, and it held the VIP tab, a pasted snapshot of top loyalty guests. It was not the source of Tuesday's answer. It is a separate external exposure, open since 2019, and the response treats it separately.

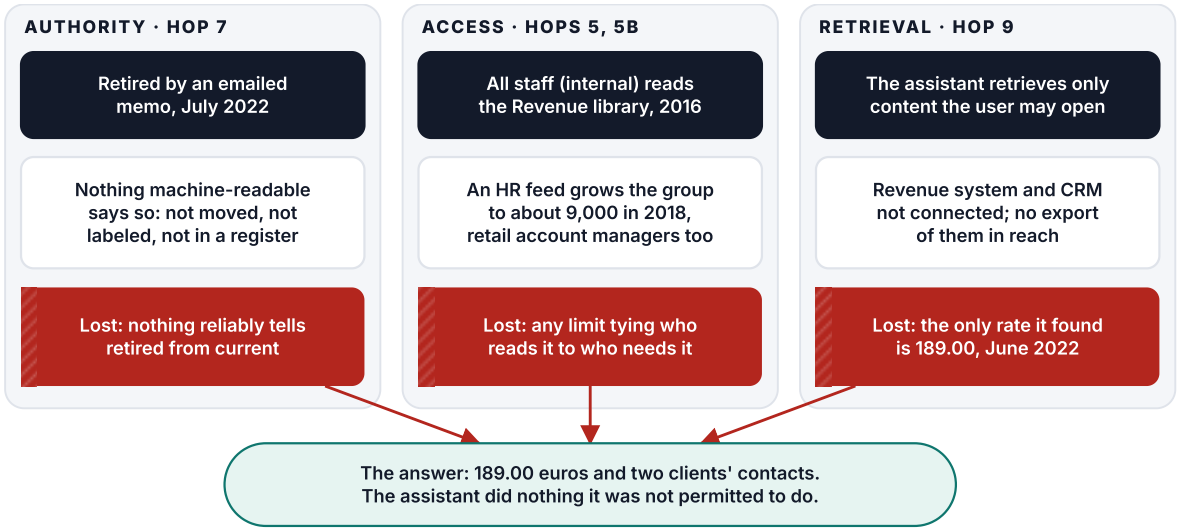
Three failures, kept apart

It is tempting to call this one failure: "the AI made a mistake." Three separate failures, made years apart by different people, made the answer possible, and each one has its own fix.

- **Authority.** The workbook was retired in July 2022 by a memo emailed to the revenue team. Nothing machine-readable says so: the file was not moved, relabeled, restricted or entered in any register (hop 7).
- **Access.** Since the 2016 migration, the custom group All staff (internal) can read the Revenue library (hop 5). Since 2018 an HR feed has made every employee a member, so membership is about 9,000 people, the retail account managers among them (hop 5b).
- **Retrieval.** The assistant sees the workbook and cannot see the revenue management system or the CRM, which are not connected; no better source was in reach (hop 9).

FIGURE P.2

Three failures, side by side: each at a different hop, under a different owner



Authority, access and retrieval failed separately, at different hops and under different owners; the assistant joined them in one answer without breaking any permission.

Author's diagram based on the book's running example; the hotel and retail group, its people and numbers are illustrative.

The assistant did nothing it was not permitted to do. Microsoft, for example, states in its documentation on data, privacy and security for its Copilot assistant, as read in November 2026, that it "only surfaces organizational data to which individual users have at least view permissions", and it asks customers to use the permission models in services such as

SharePoint "to help ensure the right users or groups have the right access to the right content". The assistant honours permissions. It does not repair them, and it cannot reliably tell that a file was retired when the retirement lives only in an email.

> *The assistant did not leak the rate sheet. The organization had already given about 9,000 people read access to it; the assistant made it easy to find and to repeat.*

What this book does with the story

Each chapter that follows takes one step of that chain forward in time; each step is a Ledger row with an owner, a reader, a cost and what was lost. In 2026 the group's search found 23 copies of the rate sheet, not counting library versions (an illustrative count). The last two chapters turn the story into work: name the authoritative source, count the copies, narrow who can open them, and retire and delete for real. The response to the incident, from the evidence kept on Thursday to the retest after the fix, is told in Chapter Fifteen.

The test this book sets is short. For one important record, can you count its copies and name who can open each one?

ERA CARD · THE RATE SHEET IN MARCH 2026

Reader. A workplace assistant retrieving, for 300 pilot users, indexed content each of them is permitted to open.

Security. Permissions set by a migration in 2016 and grown by an HR feed in 2018; a link with no sign-in from 2019; no review of either.

Cost. The workbook cost nothing to keep until someone read it; then it cost a quote, a breach assessment and the time of eight people (Chapter Fifteen).

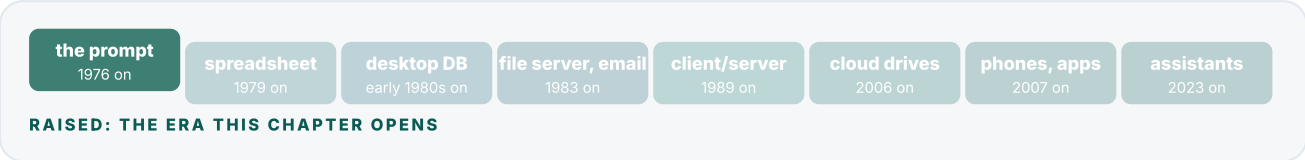
AI. The assistant answered from the only rate it found and cited it honestly; the failure was in what the organization left in reach.

Monday rule: ask the owner of your most important pricing or reporting file to show you, in writing, where the file is marked as current or retired, and who outside their team can open it.

PART I · THE FILE

CHAPTER ONE

The Command Line and the File



On 12 August 1981, at what IBM's own history calls New York's Waldorf Hotel, IBM unveiled the IBM Personal Computer. IBM's own history gives the price as 1,565 US dollars for a machine with 16 kilobytes of memory and no disk drive; a display, two diskette drives and a printer cost nearly 3,000 dollars more. Microsoft provided the operating system. Before long a manager could put a machine like it on a desk, type a record into it, copy that record onto a disk and hand it to anyone. This chapter is about what that did to the record. The decision it changes: **when anyone can create, copy and rename a record, its owner, its meaning and its standing have to be recorded somewhere other than the file, because the file will not keep them**.

From a department's system to a person's disk

Before the desk, a business record usually lived in a system that a department ran: a ledger in a bound volume, a deck of cards, a file on a mainframe. Someone managed that system and answered for it. Changing a record meant going through that department. Copying it meant a request, a print run or a tape.

The personal computer moved the record to a person. Microsoft's *MS-DOS Encyclopedia* (Microsoft Press, 1988) tells the line of operating systems that made it possible. In what it calls "the volatile microcomputer era of 1976 through 1978", Digital Research released its Control Program/Monitor, CP/M, in 1976, by the encyclopedia's dating. At Seattle Computer Products, Tim Paterson wrote an operating system for the Intel 8086 in 1980, informally christened QDOS; "four months after beginning work", the encyclopedia says, he had "a functioning 6 KB operating system, officially renamed 86-DOS". Seattle Computer Products licensed 86-DOS to Microsoft and later sold it outright; Microsoft developed it into MS-DOS. When IBM announced the Personal Computer it named three operating systems; in the encyclopedia's account only MS-DOS was available when the machine shipped, and it made its debut in August 1981. IBM sold it as the IBM Personal Computer DOS, which the trade press shortened to PC-DOS.

Who did what in those years has a large literature of its own, much of it anecdote. This book needs only the result: by the mid-1980s a business could put a general-purpose computer on a desk, and the person at that desk decided what records it held.

The prompt, the drive letter and the name

The person at that desk met the system through a prompt. The encyclopedia describes it plainly: "The standard MS-DOS prompt is A> if the system was booted from a floppy disk and C> if the system was booted from a fixed disk." The letter is a **drive letter**, a name for a disk. Microsoft's documentation still calls it a volume designator and treats it as part of a path: "C:\" in "C:\path\file".

Every record on that disk was a **file**: a named container of bytes. Its name followed a rule inherited from CP/M, which the encyclopedia states as "eight-character filenames and three-character extensions". Microsoft's file system protocol specification, [MS-FSCC], today defines an **8.3 filename** exactly: a base name of one to eight characters, optionally a period and an extension of one to three characters, only ASCII characters below 0x80, and no spaces. So the rate sheet's name could be at most eight characters and a three-character extension: `RATES.WK1`. Windows can still keep an 8.3 alias beside long names on FAT and NTFS volumes, for compatibility, where short-name creation is enabled.

Look at what the name carried. `RATES` said what the file was about, to the person who named it. `WK1` said which program could read it. Nothing in the name said whose file it was, which season it covered, whether it was the current version or a draft, or who else held a copy. The name was a label for finding the file on one disk, not a description of the record inside it.

Part of the desk	What it gave the person	What it did not record
Prompt (`A>`, `C>`)	Direct commands, no request to another department	Who typed them, or why
Drive letter	A name for a disk in this machine	Where else the same data lives
8.3 file name	A short label and a format hint	Owner, meaning, version, status
Floppy disk	A record that fits in a pocket	Who has the disk now

The fixed disk brought one more idea that matters for records. The encyclopedia explains that version 1.0 "had a single directory for all the files on a floppy disk", which became unmanageable once IBM planned the Personal Computer XT with a 10-megabyte fixed disk. MS-DOS version 2.0 introduced a hierarchical file system, modelled on Microsoft's UNIX-

based XENIX: a root directory under which users could create subdirectories. In the encyclopedia's words, "Each file in the system was identified by the directory path leading to it", and the path could not exceed 64 characters.

That sentence describes the identity most desk files still have: where they sit. Move it to another directory and its path changes. Copy it to a floppy disk and the copy has a different path and the same name. Chapter Two returns to this as a rule: a path is a locator, not an identity.

The floppy disk and the copy

The IBM machine, the encyclopedia notes, "used 5-1/4-inch disks", and MS-DOS version 1.1 added support for "double-sided, 320 KB floppy disks". A season's rate sheet fitted on one with room to spare. A floppy disk could go in a drawer, a briefcase or an envelope to a hotel.

MS-DOS made copying and renaming the most ordinary things a person did. Its command interpreter built in commonly requested functions "such as DIR and COPY, for speed and ease of use". Microsoft's documentation describes the commands that survive today in the same words. `copy` "copies one or more files from one location to another", and if you do not name the destination file, "a copy is created with the same name, modified date, and modified time as the original file". `ren` "renames files or directories".

Those two sentences are the root of this book. A copy made with `copy` has the same name, the same size and the same modified date as the original. On an MS-DOS FAT disk nothing a person would look at told the two apart; on NTFS, Microsoft documents, the copy gets a new creation time, which few people ever look at. No file system records that one file is a business copy of the other. A rename changes the path and name, the identity most people and tools use. And neither command asks who owns the record, or records that a copy now exists.

How a file system lays those bytes out on a disk is the subject of *From Bit to Pixel*, Book 1 of *Every Pixel Matters*; one paragraph is enough here. The encyclopedia credits Marc McDonald with designing the file allocation table, the FAT, for Microsoft's Stand-alone Disk BASIC, produced for NCR in 1977, as a linked list for managing disk files. In MS-DOS each file has a 32-byte directory entry that "includes a filename and extension, the file's size, the starting FAT entry, the time and date the file was created or last revised, and the file's attributes". The archive attribute is set when a written file is closed, which tells backup tools what changed. Microsoft's own overview of FAT says it "supports only read-only, hidden, system, and archive file attributes" and that "it is not possible to set permissions on files" on FAT partitions. There is no field for a person.

Records leave the department

The spreadsheet gave the person at the desk a reason to make records. Lotus 1-2-3 Release 2, Lotus staff recalled in a Computer History Museum workshop in 2004, "shipped around September 1985"; Lotus published its WKS and WK1 worksheet file formats so other programs could read them. Chapter Three tells the spreadsheet's own story. What matters here is what a worksheet on a floppy disk did to the record: it made the person who typed it its only custodian.

That is where the thread of this book begins.

THE HOP · 0 · 1987 · RESERVATIONS MANAGER'S NOTES AND THE TYPED RATE MEMO TO RATES.WK1 ON A FLOPPY DISK (LOTUS 1-2-3 RELEASE 2, MS-DOS PC)

Record	The rate sheet (hotel and retail group), hop 0, 1987
From	Reservations manager's notes and the typed rate memo
To	RATES.WK1 on a floppy disk (Lotus 1-2-3 Release 2, MS-DOS PC)
Mechanism and schedule	The reservations manager types the season's rates into a worksheet, prints it and faxes the printout to 12 hotels
Format change; what was lost or assumed	CONTEXT Worksheet cells: hotel, room type 01 to 12, date range, rate code (RACK, CORP), rate; the seasonal uplift computed by a 1-2-3 macro. **The conditions of each rate (cancellation, deposit, who qualifies for the corporate rate, and blackout dates when corporate rates do not apply, such as the harbor festival) stay in the typed memo; the sheet holds codes and numbers**
Accountable owner	Reservations manager, head office
Who can change it	The reservations manager (the only machine-readable copies are on her floppy disk and her PC; the 12 hotels hold faxed paper)
Classification; retention or erasure duty	Commercial pricing data; no retention rule
Read by (people, apps, models, agents)	Reservations staff; 12 hotels reading the faxed printout
Cost per month; basis	Per month: one PC and a software licence; per season: about two days of work (basis: staff time) (illustrative)

In 1987 the reservations manager at the group's predecessor, a hotel company with 12 hotels, typed the season's rates into a Lotus 1-2-3 Release 2 worksheet on an MS-DOS PC and saved it as `RATES.WK1` on a floppy disk. Each row held a hotel, a room type from 01 to 12, a date

range, a rate code (RACK or CORP) and a rate; a 1-2-3 macro computed the seasonal uplift. She printed the sheet and faxed the printout to the 12 hotels. The conditions of each rate stayed in the typed memo: cancellation, deposit, who qualified for the corporate rate, and the blackout dates when corporate rates did not apply, such as the harbor festival. The sheet held codes and numbers. The only machine-readable copies were on her floppy disk and her PC; the hotels held paper. Loss code CONTEXT: the notes stayed behind. Thirty-nine years later, that missing blackout is why 189.00 euros was the wrong answer.

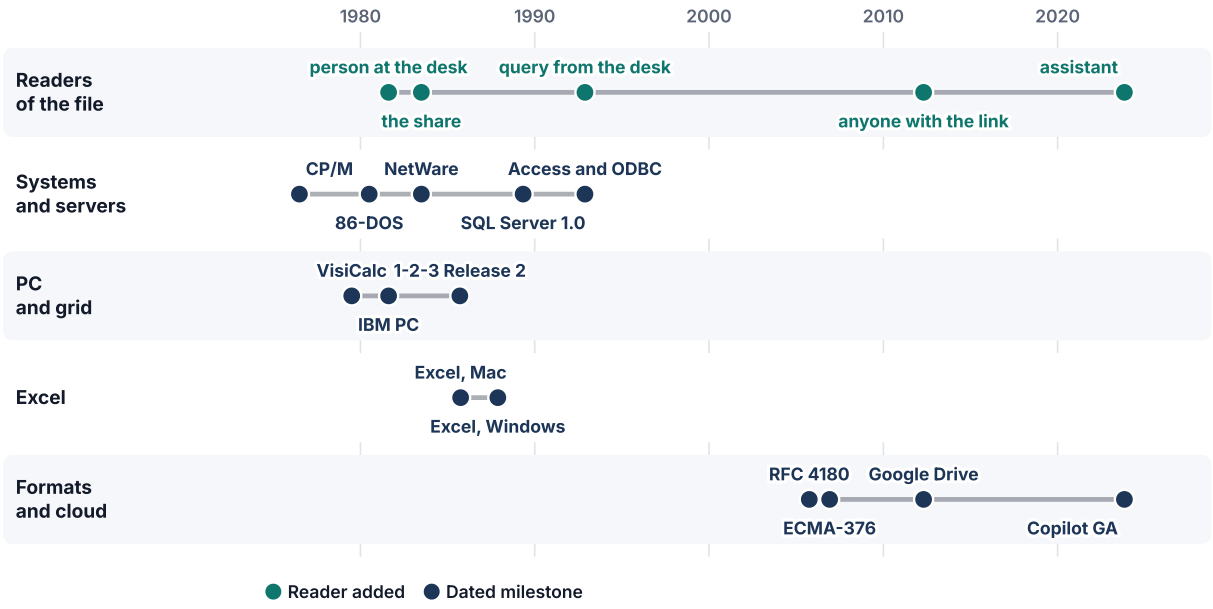
Notice what hop 0 did not do. It did not copy the record anywhere dangerous; the faxes were paper and the disk stayed on one desk. It did something quieter: it split the record in two. The numbers went into a file that could be copied, renamed and later read by machines. The rules for using those numbers stayed in a memo that no machine would ever read. Every later hop in this book copies the first half and leaves the second behind.

The desk era at a glance

Figure 1.1 places the milestones this book uses on one line, with a lane for the reader each step added. The person at the desk came first. The share on a file server added everyone with rights to a directory; the query from the desk added records pulled from servers into spreadsheets; cloud drives added anyone holding a link; and in 2023 the workplace assistant arrived, able to retrieve supported, indexed content that a user is permitted to access. Microsoft announced that Microsoft 365 Copilot was "generally available for enterprise customers worldwide" on 1 November 2023. None of the earlier readers went away.

FIGURE 1.1

The desk era with its readers: from the prompt to the assistant, each step added a reader of the file



Top lane: the reader each step added, none of whom replaced the ones before, each placed at the milestone this book uses for it, not its first instance; lower lanes: dated milestones this book uses. Points with only a year are drawn at mid-year; the assistant point is Microsoft 365 Copilot's general availability for enterprise customers, 1 November 2023. ODBC 1.0 is dated only to 1992 and drawn with Access.

Source: The MS-DOS Encyclopedia (CP/M, 86-DOS); IBM history (IBM PC); Novell (NetWare); Computer History Museum (VisiCalc; 1-2-3 Release 2 workshop transcript); Microsoft, The History of Microsoft 1985, 1987, 1989, 1992; RFC Editor (RFC 4180); Ecma International (ECMA-376); Google (Drive); Microsoft Tech Community (Copilot). Details in Appendix E.

Two lessons run through the timeline. First, each new reader inherited the files the earlier ones had made, without their context: a file saved for one person in 1987 can be in front of an assistant in 2026. Second, the tools grew better at moving files and never began recording who answers for them. Chapter Two looks at exactly what a file records, platform by platform, and at what none of them records.

WHAT HASN'T CHANGED

Microsoft's documentation for the `copy` command, as published in 2026, still says that a copy made without a new name keeps "the same name, modified date, and modified time as the original file". The command line of 1981 and the desktop of 2026 share the same blind spot: the copy looks like the original.

ERA CARD

Reader. The person at the desk, who created the record, kept its only copies and knew what it meant.

Security. Physical: whoever held the disk or sat at the PC. The FAT file system had no permissions and no field for a user.

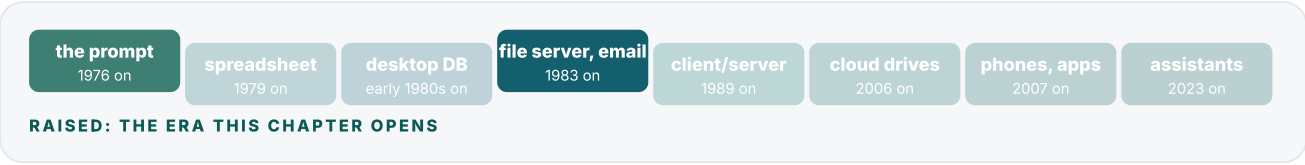
Cost. One PC, a software licence and a few days of work each season; copies cost a blank disk.

AI. None then; but the files and the habit of copying them are what an assistant reads nearly forty years later.

Monday rule: for each spreadsheet that sets a price or feeds a report, write down outside the file who owns it and where its rules and conditions live, because the file itself will not say.

CHAPTER TWO

What a File Knows and What It Forgets



On Thursday 12 March 2026 the hotel group's revenue manager opened the workbook the assistant had cited and read its file properties. They said when it was last saved: June 2022. They did not say that a memo had retired it in July 2022, that its corporate rates do not apply on festival dates, or that the public rate for those dates now lives in another system. The file knew its name, its size, its dates and who could open it. It did not know what it meant, whether it was still the record, or why it held two clients' contacts. This chapter sets out what files on each platform record and what none of them does. The decision it changes: **record each important file's authority state outside the file, in a form a machine can read, and stop treating file properties as evidence of currency or ownership**.

What a file system records

A file system keeps a small, fixed set of facts about each file. On MS-DOS, Microsoft's *MS-DOS Encyclopedia* describes the 32-byte directory entry: it "includes a filename and extension, the file's size, the starting FAT entry, the time and date the file was created or last revised, and the file's attributes". The attributes were read-only, hidden, system and archive, plus markers for a volume label and a subdirectory. The time and date fields are set when the entry is created and "updated whenever a file is written to".

Later file systems added more time. Microsoft's comparison of its file systems lists creation and last change time stamps for NTFS, exFAT, UDF and FAT32. It also shows the difference that matters for this chapter: NTFS can "track file owner" and keeps access control lists; exFAT and FAT32 do neither.

All of these fields describe the container. Size and dates are facts about bytes and writes. They answer "when did someone last save this?" They cannot answer "is this the record we use?", because the file system has no idea which other files hold the same content, or what the content means to the business.

Who may open it: trustees and access control lists

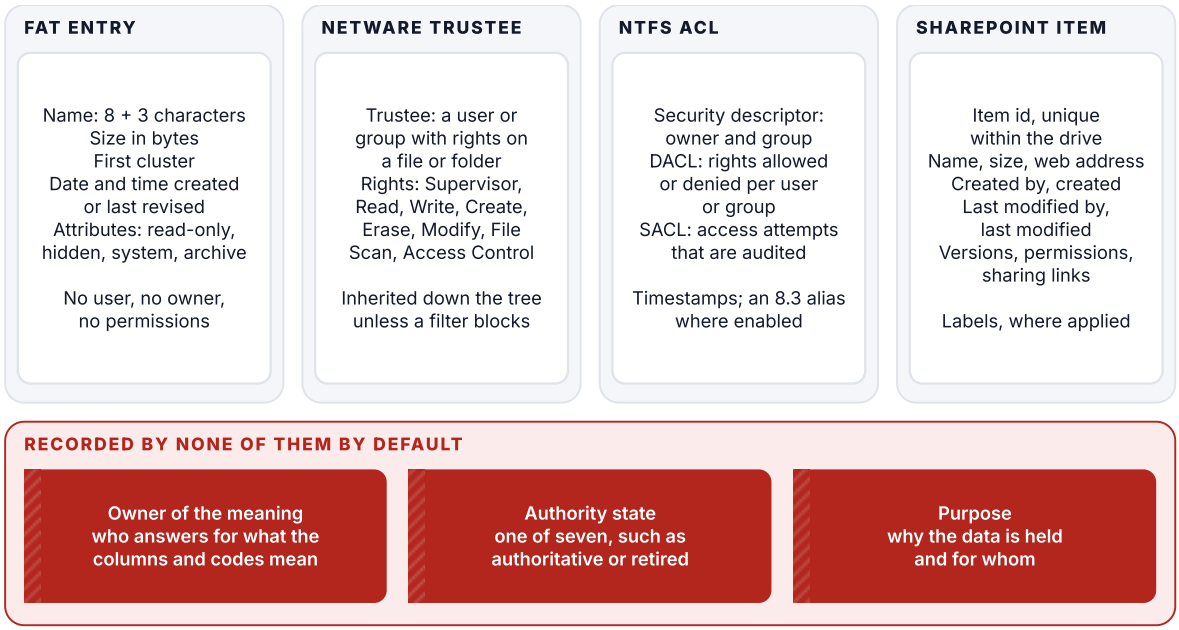
The file server added the first serious answer to "who can open this file?" On a NetWare server, which Novell announced in 1983 as its first LAN software based on file-server technology, access was granted through **trustee rights**. Novell's documentation for later NetWare releases defines a trustee as a user or object holding a rights assignment on a file or folder. The file system rights are Supervisor, Read, Write, Create, Erase, Modify, File Scan and Access Control. Rights assigned to a folder flow down to the files and folders inside it, unless an **inherited rights filter** blocks them; the Supervisor right cannot be filtered. Access Control lets a trustee change other trustees' assignments. This is the model the group's `S:\REVENUE` directory used from 1994: the reservations team and two IT staff held rights to the directory, and every file in it inherited them.

Windows NT brought access control into the file system itself. Microsoft documents that each securable object, an NTFS file among them, has a **security descriptor** holding security identifiers for its owner and primary group, a discretionary access control list (DACL) "that specifies the access rights allowed or denied to particular users or groups", and a system access control list (SACL) "that specifies the types of access attempts that generate audit records". Rights are granted to groups as easily as to people, and they too can be inherited from the folder above. The group's file server moved from NetWare trustees to Windows groups in 2006; Chapter Nine follows what that remapping did ten years later.

Both models answer one question well: who may open, change or delete this file? Neither answers whether they should, for what purpose, or whether the content is still current. And both grant through groups, whose membership changes without anyone touching the file. In 2016 one group's read grant on one library reached about 4,000 people; by 2018 it reached about 9,000, without any change to the library's own permissions.

FIGURE 2.1

What a file records, by platform, and what none of them records



Each platform added fields for location, time and access; none added, by default, a field for who answers for the meaning, which copy is authoritative or why the data is held.

Sources: The MS-DOS Encyclopedia (Microsoft Press, 1988); Microsoft KB 100108; Novell ConsoleOne 1.3 User Guide, "About NetWare Rights"; Microsoft Learn, "Security Descriptors" and "File System Functionality Comparison"; Microsoft Graph, driveItem resource. Details in Appendix E.

What the document says about itself

Office documents carry a second set of facts inside the file: **document properties**. Microsoft's support pages describe four types. Standard properties, "such as author, title, and subject", which users can set. Automatically updated properties, which users "cannot specify or change", such as dates and program statistics. Custom properties that an organization defines. And, for files in a document library, the library's own properties. In the Office Open XML formats the core set is fixed by the standard; Microsoft's .NET documentation for these core properties describes, among others, the creator ("the individual or entity that created the Package and its content"), last modified by ("the user who last modified the Package content"), created and modified dates, a revision number and a content status.

Those properties read like provenance. They mislead in three ordinary ways.

- **They travel with the copy.** The properties are inside the file, so every copy carries the original's author and last-modified-by. A copy made with the `copy` command keeps even the modified date, as Chapter One showed. Twenty copies of a workbook can all name the

same author and none of them is more authoritative for it.

- **They record a person, not an owner.** "Author" is whoever created the file, or whoever the template named. It says nothing about who answers for the content today, and nothing changes it when that person moves to another hotel or leaves.
- **They record an event, not a state.** "Last modified" says a save happened. It does not say whether the content is current, retired or a draft. Even a "status" property is free text that anyone can set to "Final".

Cloud libraries add a third set of facts, kept by the service rather than the file. Microsoft's Graph documentation for a OneDrive or SharePoint item lists, among others, a unique identifier, the name, the size, a web address, who created the item and when, who last modified it and when, its versions, its permissions and its sharing links. These are better evidence than document properties, because the service writes them. But they describe the item in that library, from the day it arrived. After a migration, they hold whatever the migration tool wrote into them, which may or may not be the values from the old file server. Before you trust a library's "created by" or "modified" values for a file older than the migration, find out what the migration tool wrote into them.

Formats are contracts

A file format is a contract between the program that writes a file and every program that reads it. The contract says how bytes become cells, formulas and formats. It says nothing about whether the numbers are right.

The desk era's contracts were often private. Lotus staff recalled, in the Computer History Museum's 2004 workshop on 1-2-3, that part of the job was "to make the worksheet file format the de facto standard for tabular information interchange", and that Lotus published its WKS and WK1 formats. Microsoft's binary Excel formats, BIFF, are now documented in its open specification [MS-XLS]: BIFF5 was "used by default to save workbooks in Microsoft Excel 5.0 for Windows and Microsoft Excel for Windows 95", and BIFF8 "to save workbooks in Microsoft Office Excel 2003 and earlier versions", which Microsoft's list of Excel file formats gives as Excel 97 to Excel 2003. The open standards came later. Ecma International published ECMA-376, Office Open XML file formats, in its first edition in December 2006, and submitted it the same month to ISO/IEC JTC 1; it was published as ISO/IEC 29500 in 2008. OpenDocument, the other open office format, was published as ISO/IEC 26300 in 2006.

Each change of contract is a hop. When the group's sheet moved from WK1 to an Excel workbook in 1994, the 1-2-3 uplift macro did not carry over and staff pasted values in its place (Chapter Three). By the time it moved to `.xlsx` in 2010, the on-premises CRM behind its live query had been switched off, so the query failed: from then the CORP tab's

client contacts stayed as they were in 2010, and its rates were typed by hand until June 2022 (Chapter Eight). A format contract tells you whether a file can be read. Only the Ledger tells you what the reading lost.

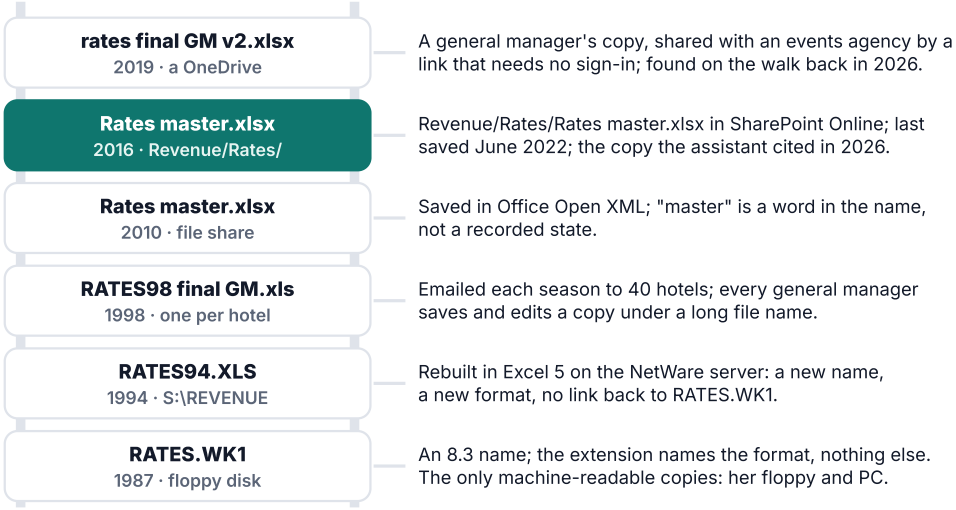
A path is a locator, not an identity

Microsoft's documentation describes a path as "a string representation of where to get the data". That is exactly what it is: directions, not a name for the record. The MS-DOS encyclopedia said in 1988 that each file "was identified by the directory path leading to it", and for most desk files that is still the only identity they have.

The group's rate sheet shows the cost. It never had a key. Its identity was a path and a name, and both kept changing.

FIGURE 2.2

Six names for one workbook: its identity was a path and a name, and both kept changing



From 1987 to 2019 the rate sheet lived under six names, counting each path as a name, each in a different place; no business identifier joined them, so the trace in 2026 had to rebuild the chain from memory, mail and file properties. The 2019 OneDrive copy is a side branch, not a step toward the answer.

Author's diagram based on the book's running example; the hotel and retail group, its people and numbers are illustrative.

Six names in 32 years, counting each path as a name: five on the causal chain and the 2019 OneDrive side copy, each in a different place. A rename looked like a new file to the search tools that indexed it; a copy looked like the original to the people who opened it. When the

revenue manager traced the 2026 answer, nothing in any of the six files pointed to the others. The chain had to be rebuilt from memory, old mail and file properties.

Some platforms do offer an identifier that survives a move. Microsoft Graph addresses a library item either "by the driveItem unique identifier" or "by file system path", and the identifier is unique within the drive. SharePoint's Document ID service gives a document an ID that, in Microsoft's words, "will normally move with the file"; it has to be activated for each site collection, and a copy gets a new ID. These help within one platform. None of them survives a copy to another platform, an email attachment or a download to a laptop. A record that matters needs an identifier recorded outside every file that holds it: a register entry that lists each known copy by path and says which one counts.

Which copy is authoritative?

That question has no answer in any field shown in Figure 2.1. It needs a separate record: the **authority record**, which names, for one kind of decision, the source to use and the state of every other known copy. This book uses seven **authority states**; the authority record holds them, and a Ledger row can carry one in one of the five optional file columns that Appendix A adds to the ten core columns.

State	What it means
Authoritative	The source of record for a named decision; an owner answers for it.
Operational copy	A controlled copy a system runs on, refreshed from the authoritative source.
Reference copy	A read-only copy kept for lookup or reporting, with its source and date stated.
Working copy	A person's draft or analysis; not for decisions outside its author's task.
Retired	Replaced; do not use for decisions. Kept only until archived or disposed.
Archived	Kept for a stated retention reason, out of everyday reach and search.
Disposed	Deleted under a rule, with a record of when and which copies.

Lineage and authority answer different questions. The Ledger shows which copy an answer used and what was lost on the way. It cannot decide which copy should have been used. That decision belongs to the authority record, made in advance by an owner. In the prologue, lineage traced the answer to `Rates master.xlsx`; only an authority record could have said, before Tuesday, that the revenue management system was the source for public rates and the workbook was retired.

Why retired must be machine-readable. The group did retire the workbook. In July 2022 a memo went by email to the revenue team. Every person who read the memo knew the file was no longer the record. But the memo was not attached to the file, and nothing on the file, or visible to someone who found it, said the workbook was retired. The file stayed in the same library, with the same permissions, the same name and no label. To the systems that read the library, from search to the assistant, a retired file and a current one looked the same. So retirement has to be done in the places machines look:

- **Location:** move the file to an archive location, or remove it.
- **Permissions:** narrow who can open it to the people who must.
- **Label:** mark it retired with a label that a policy acts on, so that search and the assistant exclude or restrict it; a label with no policy behind it is only a word.
- **Register:** enter it in the authority record with its replacement and the date.

A memo can explain the decision. It cannot carry it out.

> **OUR VIEW, AS OF NOVEMBER 2026**
>
> Most organizations will not tag every file with an authority state, and they do not need to. Start with the files that feed a price, a regulator, a model or an assistant. For each one, the register entry and the retired label are cheaper than a single wrong answer like the one in the prologue.

ERA CARD

Reader. The person at the desk, then everyone with rights to a directory on the file server.

Security. Trustee rights and access control lists granted through groups; set once, inherited down the tree, rarely reviewed.

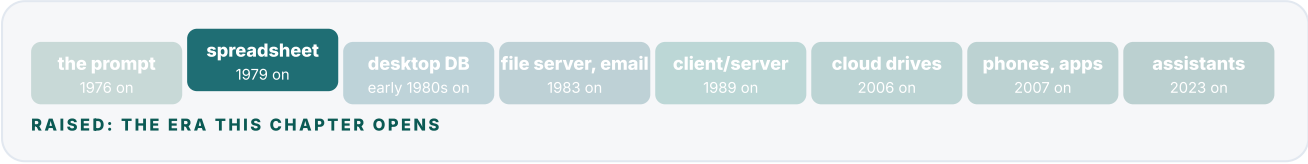
Cost. Metadata is free to keep and expensive to trust: every investigation reconstructs from properties what no field recorded.

AI. An assistant reads the same fields a person does; it can see a last-saved date, but not a retirement that lives only in a memo.

Monday rule: for your ten most important files, write an authority record that names the source for each decision and marks every other copy you know of with one of the seven states.

CHAPTER THREE

VisiCalc, Lotus 1-2-3, Excel: the Grid as a System



In 1994 the reservations manager at the hotel company's head office opened `RATES.WK1`, the Lotus 1-2-3 worksheet she had kept on a floppy disk since 1987, in Excel 5 on a Windows for Workgroups PC. The rates came across: hotels, room types, date ranges, the codes RACK and CORP. So did a small program stored in the sheet, the macro that applied each season's uplift to the base rates. Excel could still run it while the file stayed a 1-2-3 worksheet. She rebuilt the sheet and saved it as `RATES94.XLS` on the new file server, and the team pasted the uplifted rates in as plain numbers. Nobody wrote down that the rule which produced those numbers had stayed behind in the old file. From that day the rates looked the same, but the rule that would make next season's rates was gone. This chapter is about the grid, the program shape that VisiCalc, Lotus 1-2-3 and Excel made the desk's way of building systems. The decision it changes: **before you convert, move or replace a workbook that sets a price or feeds a decision, list its rules as well as its numbers, and decide where each rule will live afterwards**.

Three programs, one idea

A spreadsheet program shows a grid of cells, named by column letter and row number, and recalculates every dependent cell when one changes. The Computer History Museum's timeline credits Dan Bricklin and Bob Frankston with VisiCalc in 1979, "the program that turned the personal computer into a business machine", and describes how it let users ask "What if?" questions of their financial figures. The museum's exhibit on putting personal computers to work calls VisiCalc "the first spreadsheet for personal computers". The qualifier is the museum's, and it matters: the claim is about personal computers, not about the idea of a ruled worksheet.

Lotus 1-2-3 brought the idea to the IBM PC. The museum's timeline files its development under 1982, and its profile of Mitchell Kapor, who founded Lotus Development Corporation in 1982 with the program's co-creator Jonathan Sachs, says it was released in January 1983 and combined a spreadsheet, charting and "rudimentary database operations". In a workshop the museum recorded in May 2004, a Lotus staff member recalls that 1-2-3 Release 2 "shipped around September 1985", and another remembers publishing "the WKS and WK1 formats", the worksheet file layouts that other programs could then read. A 1987 rate sheet named `RATES.WK1` sits squarely in that world.

Microsoft's own history dates the shipment of Excel for the Macintosh to retail stores to 30 September 1985. Excel for Windows was announced on 6 October 1987 and shipped on 19 November 1987. Excel had to live beside 1-2-3, and one compatibility choice from that time is still inside every workbook on the 1900 date system, the default in Excel for Windows. Microsoft documents that when 1-2-3 was first released it treated 1900 as a leap year, which it was not, and that Multiplan and Excel made the same assumption "to use the same serial date system used by Lotus 1-2-3". Microsoft explains why the choice stays: correcting it would shift almost every stored date by one day. A grid inherits its rules from the files it must open. That is the first lesson of this chapter, and it applies to the reader as much as to the vendor.

The cell: value, formula, format

The cell is the unit of the grid, and it carries three different things that look like one. The **value** is what is stored: a number, a piece of text, a date held as a serial number. The **formula**, when there is one, is the rule that computes the value from other cells. The **format** decides what the eye sees: two decimals, a currency sign, a date pattern. A cell showing 189.00 may hold the number 189, the number 188.995, the text "189.00", or a formula whose result happens to be 189 today. On screen the four look identical. In a printout or a fax, the way the group's sheet reached its hotels in 1987, only the formatted value survives.

This is why a spreadsheet is more than a table. A database table holds values in typed columns. A worksheet can hold values, rules and presentation in the same rectangle, mixed cell by cell, with nothing to say which cells are inputs, which are results and which were typed over a result by hand. When a person pastes "values only" over a column of formulas, the sheet looks unchanged and has lost its logic. Hop 1 is a close cousin: the rule was a macro, not a formula, and it stayed in the old file.

Why did the group price from a workbook for thirty-five years, long after it had reservation and distribution systems? The answer the revenue team gave is plain: pricing rules changed faster than the reservation system's vendor could change them, and the workbook was the

place the revenue team could change them alone. That is the grid's strength and the reason it becomes a system without being designed as one. The insurer told in *From Tally to Token* reached the same place by a different road, with a 2004 workbook that a branch clerk built to correct dates and that later flowed back into the policy system.

A rate cell in the group's sheet makes the point concrete. The Type Contract, this series' seven-line description of what a value needs in order to travel without losing its meaning, shows how much the cell leaves unsaid.

TYPE CONTRACT

Type, precision and rounding. A rate per room night, held as a binary floating-point number; displayed to two decimals by format, so 188.995 shows as 189.00; whether the sheet rounds, truncates or keeps the full value is set by formulas, not by the cell.

Encoding and serialization. Stored in the workbook format of the day (WK1 in 1987, BIFF5 for Excel 5 in 1994, BIFF8 for Excel 97 from 1998, Office Open XML from 2010); in a printout, fax or CSV only the formatted text survives.

Unit or currency. Implied by the hotel's country and a currency format on the column; nothing in the value names the currency. An EU hotel priced in 1987 in a national currency; the euro was launched on 1 January 1999 and its cash followed on 1 January 2002, and no cell records which currency a rate was typed in.

Time basis. Valid for a date range typed in other cells of the same row; the conditions that suspend it, such as a festival blackout for corporate rates, live in a memo, not in the sheet.

Identity scope. Identified by position: sheet, row and column, plus the hotel, room type and rate code typed beside it; a sort or an inserted row changes the address.

Absence. An empty cell can mean "not offered", "not yet set" or "deleted by mistake"; a zero can mean free or missing.

Authority. None recorded in the cell or the file: the copy that counts is the one the reservations manager says counts.

Macros: the rule that lives beside the numbers

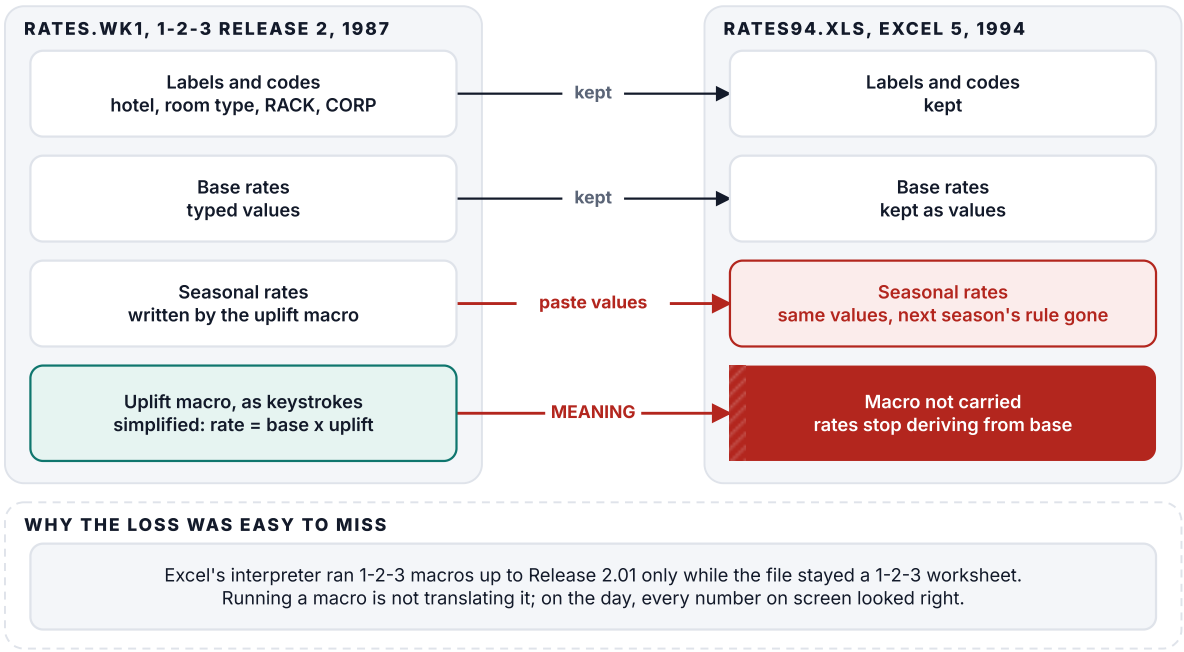
A macro is a stored sequence of commands that the program replays on request. In 1-2-3 a macro was a sequence of keystrokes stored in the worksheet under a short name such as `\U`. Jonathan Sachs, recalling the design at the museum's 2004 workshop, put it plainly: "The macro language was exactly that. It was a set of keystrokes, and it wasn't a general-purpose language." Microsoft's knowledge base article on running 1-2-3 macros in Excel describes the conventions it supported: CTRL plus the macro's letter to run it, and a macro named `\0` that runs automatically when the worksheet opens. A macro could do what a person at the keyboard could do: copy a column, multiply it by a factor, write the results over the old values. In the group's 1987 sheet, a macro applied each season's uplift to the base rates. The rate cells held numbers; the rule that made them was the macro.

Microsoft did not leave 1-2-3 users stranded. Its archived knowledge base articles for Excel 4.0 and 5.0 describe a Macro Interpreter for Lotus 1-2-3 users that could run a 1-2-3 macro in Excel, compatible "only with macros written in Lotus 1-2-3 version 2.01 or earlier", and an article for Excel 5.0 notes that file combine macros work only with the Lotus file saved in WK1 format. The Excel 97 article states that Excel runs macros stored in a 1-2-3 worksheet "if the macros were created in Lotus 1-2-3 Release 2.01", advises that Microsoft "does not recommend that you modify an existing Lotus 1-2-3 or Quattro Pro macro in Microsoft Excel", and suggests rewriting troublesome macros in Visual Basic for Applications (VBA), the macro language of new Excel macros. Excel 5.0 also shipped a Macro Translation Assistant to translate 1-2-3 macros; Microsoft notes it was not included from Excel 5.0c on. The end came with Excel 2000, which, in Microsoft's words, "does not run Lotus 1-2-3 or Quattro Pro macros".

So the conversion problem in 1994 was not that Excel could not run the rule. It was that running a rule is not the same as moving it. The interpreter ran the macro inside a 1-2-3 worksheet. Translating it was a separate, optional step that somebody had to choose, test and own. Rebuilding the sheet as an Excel workbook and pasting the results kept every number visible on the day and dropped the rule that would produce next season's numbers. Figure 3.1 shows the cells that crossed and the one that did not.

FIGURE 3.1

What the 1994 conversion kept and dropped: the numbers crossed, the rule did not



Rebuilt as an Excel workbook, the rate sheet kept its labels, codes and base rates; its seasonal rates became pasted values, and the uplift macro that produced them stayed behind.

Author's diagram of the book's illustrative thread (hop 1), based on Microsoft Knowledge Base articles 119141, 120733, 269178 and 258777 on running Lotus 1-2-3 macros in Excel.

THE HOP · 1 · 1994 · RATES.WK1 TO RATES94.XLS ON A NETWARE FILE SERVER (S:\REVENUE)

Record	The rate sheet (hotel and retail group), hop 1, 1994
From	RATES.WK1
To	RATES94.XLS on a NetWare file server (S:\REVENUE)
Mechanism and schedule	The sheet is opened in Excel 5 on Windows for Workgroups 3.11, rebuilt and saved as an Excel workbook on the head office file server
Format change; what was lost or assumed	MEANING Excel 5 workbook (BIFF5); NetWare trustee rights on the S:\REVENUE directory. **Excel could run the 1-2-3 Release 2 uplift macro while the file stayed WK1, but nobody translated it into an Excel macro (Excel 5 offered a translation assistant; the team did not use it); the rebuilt Excel workbook does not contain the macro, staff paste the values, and next season's rates can no longer be derived from the base rate**
Accountable owner	Reservations manager, head office
Who can change it	Anyone with trustee rights to S:\REVENUE (the reservations team, two IT staff)
Classification; retention or erasure duty	Commercial pricing data; no retention rule
Read by (people, apps, models, agents)	Head office reservations and finance; each hotel from a floppy copy each season (1994 to 1998)
Cost per month; basis	Per month: a share of the file server; no change in staff time (illustrative)

At hop 1 the rate sheet left the floppy disk for a shared directory. The reservations manager opened `RATES.WK1` in Excel 5 on Windows for Workgroups 3.11, rebuilt it as a workbook with a tab per hotel, one row per room type and date range, and saved it as `RATES94.XLS`, an Excel 5 workbook, in `S:\REVENUE` on the head office NetWare file server, where trustee rights let the reservations team and two IT staff change it. Excel could run the 1-2-3 Release 2 uplift macro while the file stayed WK1, but nobody translated it into an Excel macro; the rebuilt Excel workbook did not contain the macro, staff pasted the values, and next season's rates could no longer be derived from the base rate. The loss is MEANING: the numbers crossed and the definition of how they were made did not. Every later copy of the sheet, the hotel copies of 1998, the CSV files of 2003 and the master of 2010, inherits typed rates whose rule is gone. The cost on the day was nothing: a share on the server and no change in effort.

What the grid does not enforce

A managed database can provide three controls that the grid does not. A **schema** declares each column's type and rules, so a rate column refuses text and a room type refuses a value outside the list. A **transaction**, with the database's concurrency controls, groups changes so that they all happen or none do, so two people editing at once do not leave half of each other's work. An **audit trail**, where someone configures one, records who changed what, and when, inside the system. None of them is automatic just because data sits in a database, but each is available there. A worksheet in 1994 had none of them. A column was a column because a person typed a heading. When two people opened the same workbook from a file server, the second got a read-only copy, saved it under another name, and two versions of the rates existed. Nothing recorded that the rule behind a column of numbers had been left behind.

Over time the grid gained partial controls, and each one is worth knowing for what it does not do.

- **Data validation.** Microsoft documents that validation lets you "restrict the type of data or values that users enter into cells", and also that it is designed to "prevent invalid entries only when users type data directly in a cell"; when data is copied or filled, the messages do not appear. A pasted column, which is how hop 1 happened, passes untested.
- **Sheet protection.** Microsoft states that "worksheet level protection isn't intended as a security feature"; it prevents users from modifying locked cells. It stops accidents, not people who want to change a rate.
- **Version history.** Microsoft 365 keeps previous versions of a file and lets you compare and restore them, but only for files stored in OneDrive or SharePoint. A workbook on a file share or a laptop has none, and a version history records saves, not the reason for a change.
- **Macro-enabled formats.** Since the Office Open XML formats, Excel asks you to save a workbook as `.xlsm` if it contains macros you want to keep. The extension at least tells a reader that code is inside.

None of these turns a workbook into a database, and none was meant to. The question for a team is narrower: for each workbook that sets a price or feeds a decision, which of the three guarantees does the business rely on, and where does that guarantee now come from? If the answer is "the reservations manager is careful", write that down. It is a control, and it leaves with her.

The grid's edges

The grid also has a size, and the size belongs to the file format, not to the screen. Microsoft documents that beginning with Excel 2007 the worksheet is 1,048,576 rows tall by 16,384 columns wide, and that Excel 97 to 2003 workbooks, the `.xls` format, are 65,536 rows tall by 256 columns wide. When a workbook with data beyond those edges is saved in the older format, Excel's compatibility checker warns that "data beyond 65,536 rows tall by 256 (IV) columns wide will not be saved", and formula references to that region return an error. The rate sheet never came near either edge. Other workbooks did, and Chapter Four takes up a public case where the size of a file mattered.

The edge teaches something beyond capacity. A format is a contract about what a file can hold, and a conversion between formats is a place where content can be dropped with a warning a busy person dismisses. The same is true of macros, which a format may or may not carry, of formulas, which a paste may or may not keep, and of dates, which a program may read differently from the one that wrote them. Each conversion is a hop in the Ledger's sense: a move with a mechanism, an owner and a possible loss.

> **WHAT HASN'T CHANGED**
>
> Microsoft's support article on the 1900 leap year, updated in 2026, still explains that Excel treats 1900 as a leap year so that its serial dates match those of Lotus 1-2-3. A choice made for compatibility in the 1980s is still inside every workbook on the 1900 date system.

The grid as a system

Put the pieces together and the rate sheet in 1994 was a small information system. It had inputs (base rates, typed), a rule (the uplift), outputs (seasonal rates), a distribution channel (printouts, then a shared drive), users (reservations staff and hotels) and an owner by habit (the reservations manager). It had no schema, no transaction, no change log, no record of its rules apart from the rules themselves, and, after hop 1, not even that. What it did have was speed: the team could change a price in an afternoon without a vendor's release cycle.

The right response is not to wish the grid away. It is to describe it the way you would describe any system that matters: what it holds, what rules it applies, who changes it, who reads it, and what happens to each of those when it moves. The Ledger row for hop 1 does exactly that, and its MEANING code is easy to miss in a spreadsheet conversion, because the numbers on the day look right.

OUR VIEW, AS OF NOVEMBER 2026

The most valuable line in a workbook inventory is not the file size or the last modified date. It is a one-sentence answer to "which cells are rules, and where are those rules written down outside this file?" A team that can answer it for its twenty most important workbooks can convert, migrate or retire them without losing what they mean.

ERA CARD

Reader. The analyst or manager who builds and reads the sheet, and the colleagues who receive its printouts or copies, few of whom can tell a typed value from a computed one.

Security. Whoever can save the file can change any cell; sheet protection prevents accidents, not intent; macros run with the user's rights.

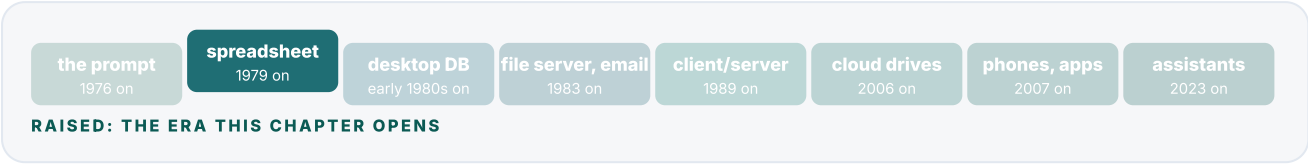
Cost. A licence and a person's time; the grid is cheap to build and expensive to convert, because its rules must be found before they can be moved.

AI. An assistant reading a workbook may see its values and formulas, but nothing in the file shows that a typed number replaced a rule, or why.

Monday rule: for every workbook that sets a price or feeds a reported number, record which cells hold rules (formulas or macros) and where those rules are documented outside the file, before anyone converts, migrates or retires it.

CHAPTER FOUR

When the Spreadsheet Is the System



On 16 January 2013 JPMorgan Chase published the report of its management task force on the losses in its Chief Investment Office, losses the firm put at approximately 5.8 billion dollars in the first half of 2012. Most of the report is about trading strategy, oversight and risk limits. Some of it is about a workbook. A value-at-risk model, a statistical estimate of how much the portfolio could lose over a set period at a stated confidence level, ran on spreadsheets fed by hand, and the report's appendix walks through what went wrong inside them. Three months later, three researchers at the University of Massachusetts Amherst published a working paper of its Political Economy Research Institute that found a range error in another famous workbook. In October 2020, Public Health England said that 15,841 positive COVID-19 cases had not been included in the daily figures because files had grown too large for the process that loaded them. This chapter is the book's one home for these cases. It reads each one by its own report, names the mechanism, and asks which control would have caught it. The decision it changes: **stop asking whether a spreadsheet is allowed and start ranking spreadsheets by what they feed, then apply the control that matches each mechanism**.

Three reports, three mechanisms

The spreadsheet failures anyone can read about are the ones somebody investigated and published; the rest stay inside the organizations that had them. Each of the three below has a primary document. The boxes keep to what those documents say, with page numbers, so a reader can check them.

SOURCED FAILURE

JPMorgan Chase, 2012: manual copy and paste, and a cell nobody checked. The *Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses* (16 January 2013) describes a new value-at-risk model for the Synthetic Credit Portfolio. When the firm's Model Review Group approved it, it noted that the calculation was being done "on spreadsheets using a manual process" and was "error prone" and "not easily scalable"; an action plan required automation, but in the report's account the process was never automated (page 105). The appendix adds that the model "operated through a series of Excel spreadsheets, which had to be completed manually, by a process of copying and pasting data from one spreadsheet to another" (page 124). The spreadsheet that ran one of the model's analytic steps "included a cell that allowed the user to switch" between two calculation methods, and a reviewer found it defaulted to the one the approval did not describe (page 128). Further review found that, in calculating relative changes in hazard rates and correlations, the spreadsheet "divided by their sum instead of their average, as the modeler had intended", an error that "likely had the effect of muting volatility by a factor of two and of lowering the VaR" (page 128). Among remedies, the report lists a review of all manual spreadsheets and "enhanced controls for key spreadsheets" in the valuation control group (page 108).

The task force does not blame a spreadsheet for the losses; it says that "direct and principal responsibility" lies with the traders who designed and implemented the trading strategy (page 7). What the workbook shows is narrower and more useful: a model that a bank relied on had no automated input, a configuration switch in a cell, and a formula error that made risk look smaller. Each of those is a mechanism with a known control.

SOURCED FAILURE

Herndon, Ash and Pollin, 2013: a range that stopped five rows short. In PERI Working Paper 322, dated 15 April 2013, Thomas Herndon, Michael Ash and Robert Pollin replicated Reinhart and Rogoff's 2010 work on public debt and growth. Reinhart and Rogoff gave them their working spreadsheet (page 5). The paper reports that "a coding error in the RR working spreadsheet entirely excludes five countries, Australia, Austria, Belgium, Canada, and Denmark, from the analysis", and a footnote gives the mechanism in one line: "RR averaged cells in lines 30 to 44 instead of lines 30 to 49" (page 7). The paper attributes a 0.3 percentage point error in the highest debt category to this spreadsheet error compounded with others (page 7). With all corrections, including selective exclusion of years and the weighting method, average real growth for countries with public debt above 90 percent of GDP was 2.2 percent, "not -0.1 percent as published" (page 1).

A range is an ordinary object in a spreadsheet and an easy one to get wrong. A formula that averages rows 30 to 44 gives a clean, plausible number. Nothing on the screen says that rows 45 to 49 hold data the author meant to include. The control is not cleverness; it is a check that counts the rows a total covers and compares them with the rows that exist.

SOURCED FAILURE

Public Health England, October 2020: files too large for the load. In its statement of 4 October 2020 (updated 5 October), Public Health England said that "15,841 cases between 25 September and 2 October were not included in the reported daily COVID-19 cases". It gave the cause as a technical issue: "some files containing positive test results exceeded the maximum file size" accepted by the process that loads these data files into central systems. It said a rapid mitigation "splits large files" and that a full end-to-end review of all systems had begun. People who tested positive received their results as normal. The statement names no software product and no row limit.

This book keeps to the statement's own words: a file grew past a size a step in the pipeline accepted, and the cases beyond that size did not arrive. The mechanism is general. Any file format, template or load job has a capacity, and a pipeline that does not compare the count it received with the count that was sent will lose the excess without an alarm.

Dates that were never dates

The fourth case has no single incident and no regulator. It is a published measurement of how often a spreadsheet's helpfulness changes scientific data. In *Genome Biology* (volume 17, article 177, 23 August 2016), Mark Ziemann, Yotam Eren and Assam El-Osta described how Excel, by default, converts some gene symbols into dates: SEPT2 (Septin 2) and MARCH1 become "2-Sep" and "1-Mar". Identifiers of another kind, such as the RIKEN accession 2310009E13, become floating-point numbers like 2.31E+13. The authors screened 35,175 supplementary Excel files from 18 journals published from 2005 to 2015, found 7,467 gene lists attached to 3,597 papers, and confirmed gene name errors in 987 files from 704 papers: about one-fifth of the papers with supplementary Excel gene lists.

The response came from the naming authority, not the software. The HUGO Gene Nomenclature Committee (HGNC) wrote in its summer 2020 newsletter that it had updated a set of gene symbols "that were being auto-converted to dates when entered in Microsoft Excel", changing symbols beginning SEPT to SEPTIN and MARCH to MARCHF during 2019. The change was published with revised nomenclature guidelines by Bruford and colleagues in *Nature Genetics* (2020, volume 52, pages 754 to 758). A community renamed its identifiers so that the tool everyone used would stop changing them.

Microsoft later added a control. Its support page "Set automatic data conversions", which applies to Excel for Microsoft 365 and Excel 2024 on Windows and Mac, lists options a user can turn off, including "Convert continuous letters and numbers to a date", "Remove leading zeros and convert to number" and "Keep the first 15 digits of long numbers and display in scientific notation if required". The page warns that data with spaces or other characters may still be treated as dates, and that the options do not directly affect data imported with Power Query. A setting on one person's copy of Excel is a personal preference, not an organizational control, unless someone decides it is.

The mechanism is automatic type conversion, here date coercion: a program decides a string is a date and stores it as one, and the original text is gone. Its numeric twin turns a room type 04 into 4, and a date such as 04/05 is read as day-month or month-day depending on regional settings; Chapter Seven takes both up for CSV files. The control is to import identifiers as text with declared column types, and to test a sample of known awkward values after every load.

Macros: code inside a document

A macro is a program that travels inside a file. That made it powerful and, as its own designers knew, costly. At the Computer History Museum's 2004 workshop on Lotus 1-2-3, Jonathan Sachs recalled warning that a keystroke macro language would be "really quick and dirty and powerful, but that there was going to be a price to be paid down the line". Chapter Three followed the price for one rate sheet: a rule that did not survive a conversion. The other price is security.

Microsoft's developer documentation for Office 2000 describes the problem as it then stood. Macro viruses used application macro languages "such as VBA, to produce viruses that are associated with documents"; one of the first, the Word Concept (also called Prank) virus, forced documents to be saved as templates. Excel macro viruses copied themselves to `Personal.xls` or some hidden location and replicated to other workbooks as they were opened or created, and, in the document's words, a macro virus "is capable of doing anything that can be performed from VBA, such as renaming and deleting files". Office 97 could warn before opening a document with macros; Office 2000 added support for anti-virus software and digital signatures on VBA projects.

The warning dialog moved the decision to the person least able to make it: a user who wants the file to work. In 2022 Microsoft changed the default. Its Microsoft Learn page "Macros from the internet are blocked by default in Office" (dated 17 July 2026 when read for this book) opens with the reason: "VBA macros are a common way for malicious actors to gain access to deploy malware and ransomware." Office applications on Windows now block macros in files that carry the Mark of the Web, the tag Windows adds to files from the

internet such as email attachments, and show a banner without an "Enable content" button. The change started rolling out to Current Channel (Preview) with Version 2203 on 12 April 2022 and to Current Channel with Version 2206 on 27 July 2022; the Semi-Annual Enterprise Channel received it on 10 January 2023. Microsoft names the ways to allow trusted macros: trusted locations, trusted sites, and macros signed by a trusted publisher. It warns that macros in trusted locations are not blocked, so those locations must be managed carefully.

Notice what the default does and does not cover. It is about files from the internet. A macro-enabled workbook opened from an internal file share in the Local intranet zone, or from a trusted location, is outside it. Many business macros live exactly there. The control for them is not a default; it is an inventory of macro-enabled files that feed decisions, an owner for each, and a rule that code which sets a price or a reported number is signed or reviewed like other code.

End-user computing as a risk class

"End-user computing" is the name risk managers give to applications built by business staff outside IT's development process: spreadsheets, desktop databases, scripts. The term matters because it moves the question from the tool to the use. A workbook that lists team birthdays and a workbook that sets an insurer's reserves are the same kind of file and completely different kinds of risk.

The insurer, founded in 1898, keeps its reserving workbook in this class. Each month an analyst pastes an extract of paid and reported claims from the claims system into the workbook. The reserving actuary arranges the claims by the year they occurred and the years since, selects development factors, and produces an estimate of claims incurred but not yet reported. That estimate becomes part of the reserve in the financial statements. The workbook is small and old, and it carries one of the larger numbers on the balance sheet.

If the insurer's reserving actuary practises in the United States, the governing text is Actuarial Standard of Practice No. 56, *Modeling*, adopted by the Actuarial Standards Board in December 2019 and effective for work performed on or after 1 October 2020. It defines a model as "a simplified representation of relationships among real world variables, entities, or events" and model risk as the risk of adverse consequences from reliance on a model that does not adequately represent what is being modeled (sections 2.8 and 2.9). Its scope names reserving models among those to which it may apply (section 1.2). It asks the actuary to perform sufficient testing that the model reasonably represents what it models (section 3.6.1), to validate its output (section 3.6.2), to use "reasonable governance and controls" to mitigate model risk (section 3.6.4) and to consider documentation another qualified actuary could assess (section 3.7). Its informational appendix gives examples of governance and

controls, including limits on access to use and modify a model and a model change management process. Its appendix also notes that models range from simple spreadsheets to complex capital models.

Bank regulators approach the same question from a different side. The Federal Reserve's SR 11-7 (4 April 2011) said that "user-developed applications, such as spreadsheets or ad hoc database applications used to generate quantitative estimates, are particularly prone to model risk". On 17 April 2026 the Federal Reserve, the OCC and the FDIC issued SR 26-2, which "supersedes and replaces" SR 11-7. Its definition of a model is a complex quantitative method that applies statistical, economic or financial theories to produce quantitative estimates, and it says that the term "excludes simple arithmetic calculations, such as those found within spreadsheets". SR 26-2 is banking guidance, expected to be most relevant to banking organizations with over 30 billion dollars in assets; it does not bind an insurer. Neither text decides the question by file type. This book reads them the same way: the file format does not make something a model; the method and the reliance do. The insurer's reserving workbook applies actuarial statistical methods to produce an estimate that people rely on, so it is a model in substance. The group's rate sheet, which multiplied base rates by a factor, is arithmetic; it still needs an owner and a Ledger row, but not a model validation.

The Ledger gives the reserving workbook the same ten columns as any hop. The row below is illustrative; it shows what a team writes down, not a real company's controls.

Ledger column	The insurer's reserving workbook (illustrative)
Record	The insurer's reserving workbook, monthly claims development
From	Claims system: monthly extract of paid and reported claims, as a CSV file
To	Reserving workbook, data tab feeding the development triangles
Mechanism and schedule	An analyst pastes the extract into a fixed range each month; the reserving actuary selects factors; the estimate is posted quarterly
Format change; what was lost or assumed	COMPLETENESS Text becomes cells in a fixed paste range; rows beyond the range fall outside the triangle formulas unless someone extends them, and no count is compared with the extract
Accountable owner	The reserving actuary, named in the insurer's model inventory
Who can change it	Two analysts and the actuary (folder rights); no record of which cells each changed
Classification; retention or erasure duty	Financial reporting record and model under ASOP No. 56; claim amounts and dates, no claimant names; retained under the financial records schedule
Read by (people, apps, models, agents)	The actuary, finance (reserve posting), internal and external auditors; any assistant licensed to a user with folder rights
Cost per month; basis	About two days of analyst time and a half day of review (illustrative)

Writing the row exposes the control gap at once: nothing compares the number of claims pasted with the number extracted. It is a cousin of the PHE mechanism, a capacity nobody checks, and the first fix is a count reconciliation cell that the actuary signs, followed by checks of totals and keys against the extract.

Rank, do not ban

The obvious response to these cases is to ban spreadsheets from anything important. It fails, and the reason is in Chapter Three: the grid exists because it lets the people who own a decision change its rules without waiting for a vendor. A ban moves the work into files nobody declares. The better answer is to rank. A spreadsheet earns controls when it feeds a system of record, a regulator, a price, a model or an assistant. Everything else is left alone, on purpose and in writing.

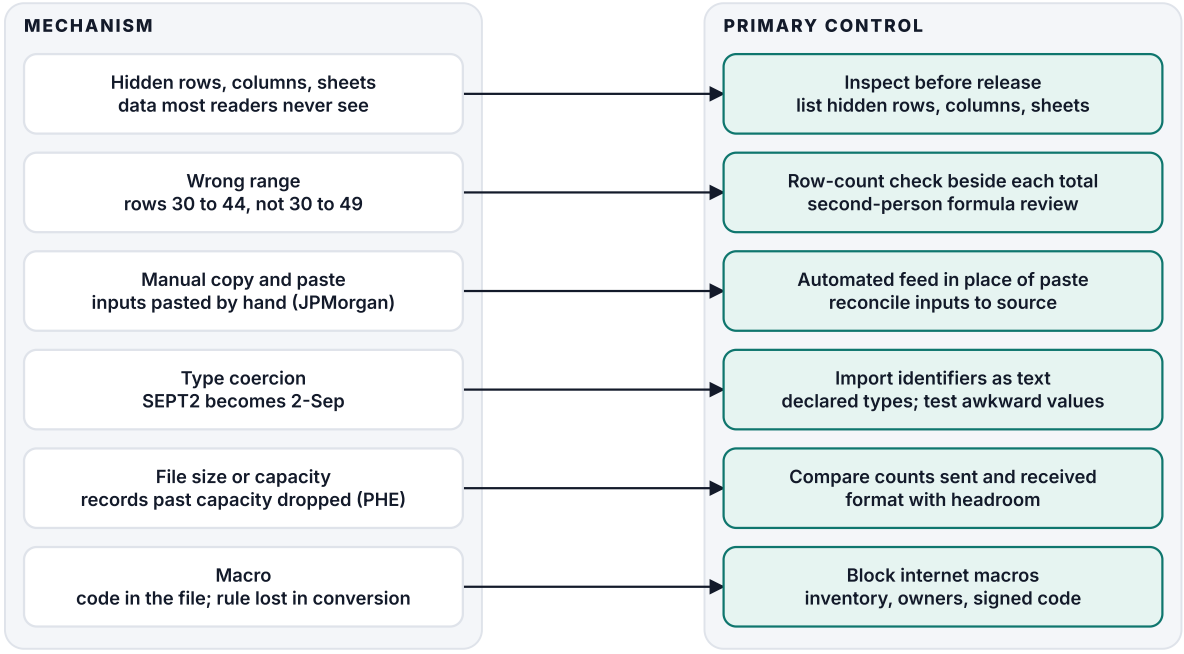
Controls then follow mechanisms, not tools. Table 4.1 lists six mechanisms that this chapter and the previous one show or point to, a public case where one exists, and the primary control for each.

Mechanism	Public case	What happens	Primary control
Hidden rows, columns or sheets	None in this chapter (Chapter Twelve has one)	Data travels with the file that most readers never see	Inspect before release; Microsoft's Document Inspector lists hidden rows, columns and worksheets, and warns that removing them can change results
Wrong range	Herndon, Ash and Pollin (2013)	A total or average stops short of the data	Row-count check beside each total; second-person review of formulas that feed a published number
Manual copy and paste	JPMorgan Task Force (2013)	Inputs drift, are pasted over formulas or land in the wrong place; a switch or formula in a cell goes unchecked	Automated feed or query in place of paste; input reconciliation to the source; review of key formulas and settings
Type coercion (dates, numbers)	Ziemann, Eren and El-Osta (2016)	Identifiers become dates or numbers; the original text is lost	Import identifiers as text with declared types; test known awkward values after load
File size or capacity	Public Health England (2020)	Records beyond a capacity are dropped	Compare counts sent and received at every hop; choose a format with headroom
Macro	Macro viruses (Microsoft, Office 2000 documentation)	Code runs with the user's rights; rules vanish in conversion	Block internet macros; inventory macro-enabled files that feed decisions; signed or reviewed code

Figure 4.1 draws the same map. None of the controls is new. What is usually missing is the decision about which files get them.

FIGURE 4.1

Mechanism to control: each way a spreadsheet fails has a primary control



Six mechanisms from the public cases and the thread, each matched to the primary control that reduces it; none of the controls is new.

Author's diagram based on the JPMorgan Task Force report (2013), PERI Working Paper 322 (2013), the Public Health England statement of 4 October 2020, Ziemann, Eren and El-Osta (2016) and Microsoft documentation on the Document Inspector, automatic data conversion and internet macros.

Leave it alone when all of these are true. Write the list down, so that leaving a file alone is a decision and not an accident:

- It feeds no system of record, regulatory return, published number, price or model, and nothing in it would mislead or expose anyone if an assistant repeated it.
- Its output is used only by the person who built it, or by a team that can rebuild it from a system of record in an afternoon.
- It holds no personal data beyond business contact details already visible to everyone who can open it.
- It is not shared outside the team by link or attachment.
- It has a named owner and an end date, or it lives in a personal workspace that is cleared when its owner leaves.

When any line is false, the file needs a Ledger row, an owner, the control for each mechanism it uses, and an exception record if the business chooses to accept a gap. An exception has an owner and a review date; without them it is a permanent waiver.

OUR VIEW, AS OF NOVEMBER 2026

A list of the twenty spreadsheets that feed prices, reported numbers, models and assistants is worth more than a policy covering all spreadsheets. The check-type controls in Table 4.1 cost about an hour per file to add; replacing a paste with a feed costs more. The expensive part is finding the files, and that is the work of Chapter Fourteen.

ERA CARD

Reader. The analyst, the actuary, the modeler and the people who rely on their output without seeing the workbook: managers, auditors, regulators and the public.

Security. Macros are code with the user's rights; the default block covers files from the internet, not the internal shares where business macros live.

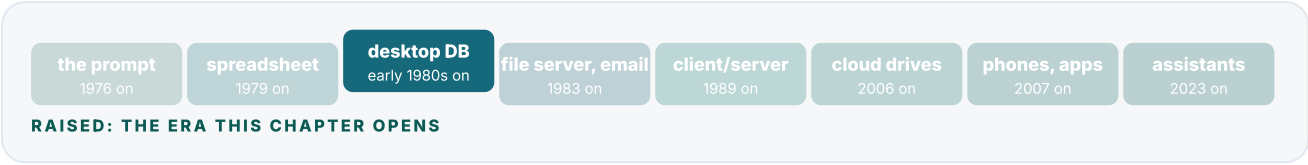
Cost. Controls cost hours per file; the failures in this chapter cost a misstated risk measure, a contested paper and delayed public health data.

AI. An assistant summarizing a workbook reports its totals as found; it cannot see that a range stops short or that a column was coerced.

Monday rule: list every spreadsheet that feeds a price, a reported number, a model or an assistant, and for each one add the control that matches its mechanism, starting with a count check between source and sheet.

CHAPTER FIVE

dBase and Access: the Desktop Database



In 1996 the sales coordinator at the harbor hotel, one of the company's hotels in the European Union, built a small database to price group bookings. Each season head office sent the hotel a copy of the rate sheet on a floppy disk, and she saved it as `C:\RATES\RATES94.XLS` on the hotel's PC. Instead of retyping its rates, she linked the workbook into an Access 95 database, so that her group quotes always read the rates from the file. It worked for two years. Then the sheet began to arrive by email, under a new name and in the new Excel 97 format, last season's file was cleared from `C:\RATES`, the link failed on open, and nobody at the hotel knew how to relink it. The staff did the sensible thing under pressure: they imported the last table they could reach, last season's sheet from the floppy disk it came on, and kept quoting from it. For two seasons groups were priced from rates for a season that had ended. Nobody owned the database, so nobody owned the problem. This chapter is about the desktop database: tables, keys and forms on the desk, the shared file on a server, and the link that points at a path. The decision it changes: **every desktop database in use gets a named owner and a list of the files and paths it links to, and a link that breaks is fixed at its source, not replaced by a frozen copy**.

The database comes to the desk

The spreadsheet put calculation on the desk. The desktop database put structure there: tables with named fields, records that could be sorted and searched, and screens for entering them. The Computer History Museum's corporate history of Ashton-Tate traces dBASE II to a program called Vulcan, written by C. Wayne Ratliff. In August 1980 George Tate and Hal Lashlee formed Software Plus, the predecessor of Ashton-Tate; a user of Vulcan suggested they distribute it, Ratliff agreed, and because the name had trademark issues the product was renamed dBASE II. The same history records Ashton-Tate announcing a multi-user version of dBASE II in February 1984: the desktop database was already being shared.

Microsoft's own history dates the next landmark precisely. On 16 November 1992, at COMDEX Fall in Las Vegas, Microsoft announced the availability of Access for Windows. Access put tables, queries, forms and reports into one file, with a graphical designer for each. A person who knew the business and not much about databases could build a working application in days.

Three ideas made the desktop database more than a list. A **table** holds records of one kind, each with the same fields. A **key** identifies each record: Microsoft defines a primary key as "a field or set of fields with values that are unique throughout a table", never empty, and a foreign key as "another table's primary key" stored in a related table to point back to it. A **form** is a screen for entering and editing records, which can check entries as they are typed. Together they give a desk application something the grid did not have: an identity for each row that does not depend on its position, and rules that apply to every record entered through the form.

That is real progress, and it is also why a desktop database is easy to mistake for a system of record. It has keys, so it looks authoritative. It has forms, so it looks controlled. It does not, by itself, have an owner, a backup schedule, a retention rule or a record of who changed which row. Those come from the organization around the file, or they do not come at all.

One file, many users

An Access database of the 1990s was a single file, with the extension `.mdb`, holding both the data and the application. The simplest way to share it was to put it in a folder on a file server and let everyone open it. Microsoft still describes this as the option that "is the simplest and has the fewest requirements", and lists its costs: performance can suffer "because all database objects are sent across the network", and it is "less secure than other sharing methods because each user has a full copy of the database file".

The better pattern is to split the file in two. Microsoft describes the result as "a back-end database that contains the data tables" and "a front-end database that contains all the other database objects such as queries, forms, and reports". The back end sits in a shared network location; each user gets a front end whose tables are links to it. Microsoft lists the benefits: performance improves because "only the data is sent across the network", the back end is less likely to become corrupted, and on an NTFS file system the back end can be protected with NTFS security. It also gives one instruction that matters for this chapter: when you choose the back end's location, specify "the UNC path of the location instead of using a mapped drive letter". A UNC path names the server and share, such as `\\server\share\claims.mdb`; a mapped drive letter such as `S:` means whatever each PC has mapped it to.

Size has a ceiling too. Microsoft's Access specifications page, read on 20 November 2026 for current versions of Access, gives the total size of an Access database (`.accdb` or `.mdb`) as "2 gigabytes, minus the space needed for system objects", and notes that you can work around the limit by linking to tables in other Access databases, each of which can be as large as 2 GB. The same page lists 255 concurrent users. Both numbers are ceilings of the file format, not promises of performance, and neither says anything about whether the file has an owner.

The table that is not there

A linked table is the desktop database's most useful trick and its quietest risk. It looks like a table in the database window, but it holds no rows. It holds a description of where the rows are: a file, a path, a sheet or table name. Microsoft's guide to managing linked tables notes that deleting a linked table "only removes the information used to open the table in the data source"; the data stays where it was. Its Linked Table Manager exists to "refresh a data source and its linked tables to ensure that the data source is accessible" and to "relink a data source to change its location". Both features are there because sources move.

A link to a spreadsheet has an extra property. Microsoft documents that when you link to an Excel worksheet or named range, Access "creates a new table that is linked to the source cells", that changes made to the source cells in Excel appear in the linked table, and that you cannot edit the contents of that table in Access: to change data, you change the workbook. That rule is later than the harbor hotel's database. Microsoft removed the ability to add or edit data in a linked Excel worksheet from Access 2002 in an update of 18 October 2005, and from Access 2003 with Office 2003 Service Pack 2; in 1996 an edit in an Access 95 linked table could still change the workbook. The harbor hotel chose to change rates only in the workbook, so the link worked as a window onto a file head office maintained, though nothing stopped anyone at the hotel from editing that local copy.

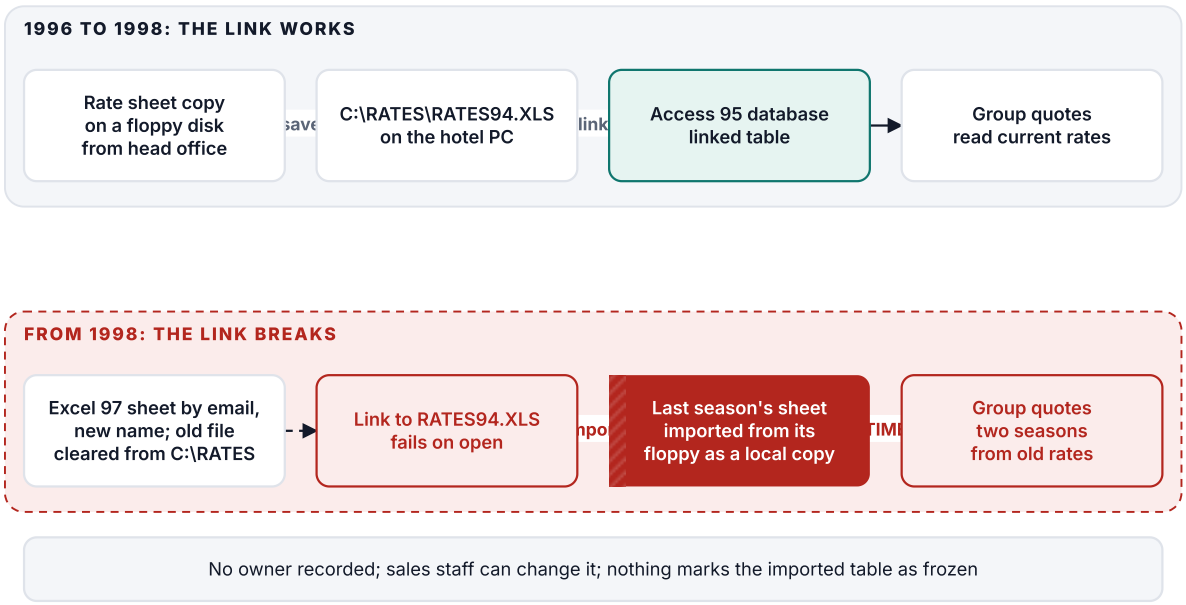
Could Access 95 do this in 1996? Microsoft's knowledge base article 149051, for Access for Windows 95 version 7.0, quotes its own help topic: "You can import or link data from a Microsoft Excel (versions 2.x, 3.0, 4.0, 5.0, and 7.0) spreadsheet." The article corrects only the first item, saying Access for Windows 95 cannot import an Excel 2.x spreadsheet. An Excel 5 workbook such as `RATES94.XLS` was within what Access 95 could link.

The weakness is the path. A link to `C:\RATES\RATES94.XLS` is a promise that a file with exactly that name will be at exactly that place each time the database opens. Rename the file, save the new season's sheet somewhere else, or move the database to another PC, and the promise breaks. What happens next depends on the person at the desk. Relinking takes a minute if someone knows how and knows which file is current. Importing the last reachable

table takes a minute too, and it turns a live window into a frozen copy that looks exactly like the live one. Figure 5.1 shows the harbor hotel's database before and after the day the link broke.

FIGURE 5.1

The harbor hotel's linked database, and the day the link broke



From 1996 the group-bookings database read rates through a link to the hotel's local copy of the workbook; in 1998 the sheet arrived by email under a new name and in the Excel 97 format, last season's file was cleared away, the link failed and an imported copy froze the rates for two seasons.

Author's diagram of the book's illustrative thread (hop 1b); linking Excel 5.0 workbooks in Access for Windows 95 per Microsoft Knowledge Base article 149051.

THE HOP · 1B · 1996 · THE HARBOR HOTEL'S COPY OF RATES94.XLS (SENT EACH SEASON ON A FLOPPY DISK) TO THE HARBOR HOTEL'S ACCESS GROUP-BOOKINGS DATABASE (LINKED TABLE)

Record	The harbor hotel's group-bookings database (side branch of the rate sheet), hop 1b, 1996
From	The harbor hotel's copy of RATES94.XLS (sent each season on a floppy disk)
To	The harbor hotel's Access group-bookings database (linked table)
Mechanism and schedule	The harbor hotel's sales coordinator links the hotel's local copy of the workbook into an Access 95 database that prices group bookings
Format change; what was lost or assumed	TIME Access .mdb with a linked table pointing at C:\RATES\RATES94.XLS on the hotel PC. **When the sheet arrives by email in 1998 under a new name, as an Excel 97 workbook, a version not in the list Access 95's help gives for linking, and last season's file is cleared from C:\RATES, the link fails on open and nobody relinks it; staff import last season's sheet from its floppy disk as a local copy and quote group rates from it for two seasons**
Accountable owner	None recorded (built by the harbor hotel's sales coordinator)
Who can change it	The harbor hotel's sales staff
Classification; retention or erasure duty	Commercial pricing data; no retention rule
Read by (people, apps, models, agents)	Group sales staff at one hotel
Cost per month; basis	Per month: near zero; one-off: two seasons of mispriced group quotes (not monthly) (illustrative)

At hop 1b the rate sheet grew a side branch at one hotel. The harbor hotel's sales coordinator linked the hotel's local copy of the workbook, `C:\RATES\RATES94.XLS` on the hotel PC, sent from head office each season on a floppy disk, into an Access 95 database that priced group bookings. There was no network link to head office in 1996; the local copy was the only one the hotel could reach. When the sheet arrived by email in 1998 under a new name, as an Excel 97 workbook, a version not in the list Access 95's help gives for linking, the old file was cleared away and the link failed on open. Staff imported the last table they could reach, last season's sheet from its floppy disk, as a local copy and quoted group rates from it for two seasons. The loss is TIME: the table kept a season's rates and lost the fact that the season

had ended. No owner was recorded: the coordinator built it, but nobody was named to answer for it; the hotel's sales staff could change the database; it cost nearly nothing to build and two seasons of mispriced groups to run.

The temporary system with no owner

The insurer, founded in 1898, has its own Access file, and in 2026 it is still open on someone's screen, converted to newer file formats along the way. In 1996 a claims supervisor built a tracker in Access for litigated claims: each claim's number from the claims system, the court, the next deadline and the outside counsel. The claims system had no place for those fields, and the supervisor was told a release planned for the following year would add them. She split the database properly, with a back end on the claims department's share and a front end on each handler's PC, and she linked the claim numbers to a nightly extract from the claims system.

The release came and covered most of the fields. It did not cover the deadline reminders, which the tracker's forms showed in red each morning, so handlers kept using it. The supervisor retired in 2009. The share was migrated twice; the back end moved with it and was converted when Access 2013 could no longer open Access 97 databases, and each time a colleague relinked the front ends from a note on the department's wiki. The extract it links to now comes from a reporting database, not the claims system, because the old extract job was switched off. Nobody is named as owner. The tracker holds the names of claimants and opposing parties, and no retention rule has ever been applied to it. These details are illustrative, but the pattern is common enough to have a name in this book: the temporary system with no owner.

The fix is not to delete it on Monday. Handlers rely on its deadline reminders, and a missed court deadline costs more than any audit finding. The fix is to treat it as what it has become. It gets a Ledger row and an owner. Its links are listed, with the source each one now reads. Its personal data gets a retention rule. Then there is a plan to move the reminders into the claims system and retire the tracker properly, with its data archived or deleted according to that rule. Until then, its back end belongs on the inventory of files to check before an assistant is connected to that share, because it holds exactly the kind of personal data an assistant should not surface by accident.

> **WHAT HASN'T CHANGED**
>
> Microsoft's current instructions for splitting an Access database still tell you to give the back end's location as a UNC path rather than a mapped drive letter, and its Linked Table Manager still exists to relink tables whose source has moved. Links to paths were fragile in 1996 and they are fragile now.

OUR VIEW, AS OF NOVEMBER 2026

Desktop databases rarely fail loudly. They fail by freezing: a link breaks, someone imports a copy, and the copy goes on looking current. An inventory that records, for each desktop database, what it links to and when each link was last refreshed will find more stale data than any scan of file dates.

ERA CARD

Reader. The staff member who uses the forms every day, often in a different role from the person who built the database, and rarely aware of what its tables link to.

Security. File share permissions on the back end, or none on a local PC; anyone who can open the file can usually read every table.

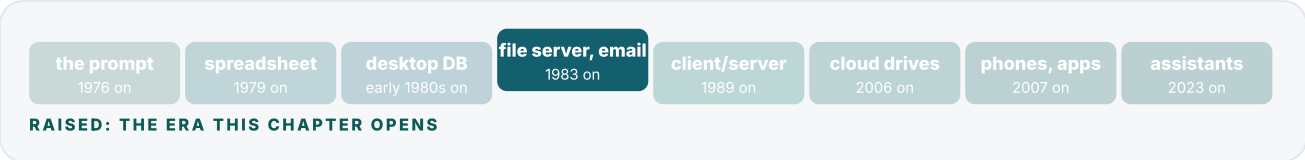
Cost. A licence and a few days to build; the cost appears later, as frozen copies, relinking after every migration and personal data with no retention rule.

AI. If an assistant or its connector can reach the file or an export of it, it cannot tell that an imported table stopped updating years ago.

Monday rule: list every Access or other desktop database file on your shares that people opened in the last year, and for each one record an owner, the files and paths it links to, and the date each link last refreshed.

CHAPTER SIX

Shared Drives, Email and the Copy as Default



In 1998 the reservations manager at the hotel group's head office stopped mailing floppy disks. The rate sheet sat on the file server, in the S:\REVENUE folder, and the group now ran 40 hotels. That season she attached the workbook to an email and sent it to every general manager. Each one saved the attachment, changed what the hotel needed, and gave the file a name that meant something to them. Within a week there were 41 rate sheets: one on the server and one on each hotel's PC. Nothing in any of them said which was the rate sheet. The practice continued each season until a revenue system replaced the workbook in 2022. This chapter covers the local network, the mapped drive and the attachment, the three tools that made the copy the default way to share. The decision it changes: **when you send a file instead of a location, record who now holds a copy and who answers for it, or accept that nobody does**.

The server down the hall

A local area network joins the PCs in one building so that they can share a printer, a disk and, later, mail. The piece that mattered for records was the file server: one machine that held folders for everyone else. Novell's own press history dates NetWare, its file server software for PC networks, to 1983. Microsoft dates the first Windows NT release, Windows NT 3.1 with its server edition, to 27 July 1993, and Windows servers took over many of the same jobs through the following decade. The hotel group followed that path: a NetWare server from 1994, its groups remapped onto a Windows file server in 2006.

On a Windows network, the protocol underneath is Server Message Block. Microsoft's documentation describes SMB as "a network file sharing protocol" with which an application "can access files or other resources at a remote server", reading, creating and updating them there. The user rarely sees the protocol. The user sees a drive letter. Microsoft's description is plain: when you "map a network drive or access a UNC path like \\server\share", the SMB client on your PC makes the connection. At head office, S: was the revenue share. In 1998 it

still sat on the NetWare server, which used Novell's own file protocol until the 2006 move to Windows; the drive letter looked the same either way. To the reservations manager it looked like a second hard disk.

That drive letter changed how people thought about a file. A document on S: was in a shared place, with permissions set by an administrator (Chapter Two covers what those rights recorded and what they left out). It could be read by many people without being copied. For a few years, the server was the closest the desk came to a single copy. The trouble began at its edge. The server served the people on its network, and the hotels were not on it. In 1998 a hotel was a separate office with its own PC, a modem and a mailbox. To reach it, the file had to leave the server.

The attachment as distribution

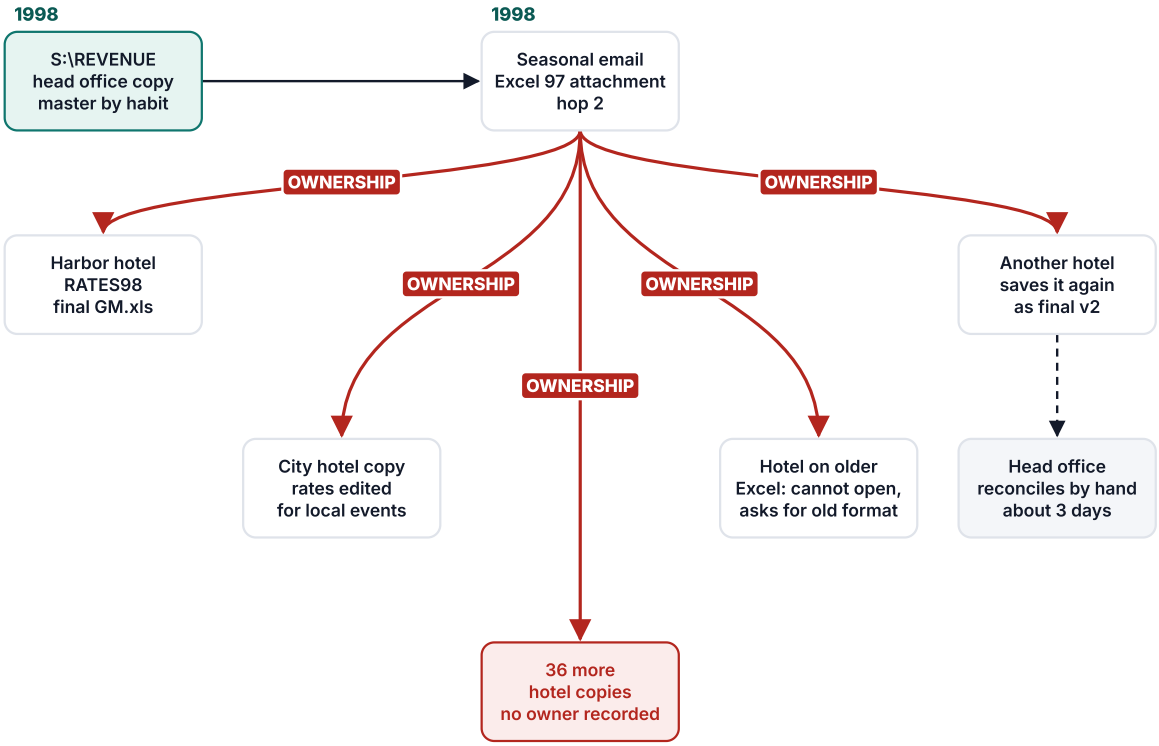
Electronic mail had carried text for years. The standard that let it carry files in a common way is MIME, Multipurpose Internet Mail Extensions, first published in June 1992 as RFC 1341. Its current first part, RFC 2045, published in November 1996 as a Draft Standard, defines the format of message bodies, including bodies that are not plain text. With MIME and the mail programs built on it, a workbook could ride inside a message. Sending a file became as easy as sending a note.

An attachment is not a pointer to the file. It is a copy, one per recipient, and each recipient's mail program can save it as another copy. A message to 40 general managers puts a copy in each of 40 mailboxes before anyone opens one (however the mail system stores them), plus the sent item in the sender's mailbox. Each copy then lives on its own terms. A general manager can forward it to the front office, save it under a new name, change two rates for a local event and send it back to head office with a question. Nothing in the message records what happened next. The sender can see that it left. The sender cannot see where it went.

Figure 6.1 draws one season of this. The copy on S:\REVENUE stays the master, but only by habit. A few hotel copies are named; the rest are a number. Each arrow carries the same loss: a copy with no accountable owner.

FIGURE 6.1

One workbook, 40 copies: each season's email makes a copy per hotel, and no hotel copy records an owner



Red chips: the loss code on each copy (Chapter Two). Dashed arrow: a copy sent back by email.

Hop 2 in 1998: the S: copy stays the master by habit; every hotel that can open it saves and edits its own copy, and each arrow carries the OWNERSHIP loss.

Author's diagram based on the book's running example (hop 2); the hotel group, its copies and the reconciliation time are illustrative.

Two things make the attachment worse than the server for records. The first is that mail keeps copies on its own schedule. A message sits in the sender's sent items, the recipients' inboxes and any folder a person moves it to, and it stays until someone deletes it or a retention rule removes it. The second is that a recipient can be anyone the sender types. The server had a list of people allowed to read S:\REVENUE. The mailbox had no such list for an attachment once it was sent. When personal data later went into the workbook (Chapter Eight), every season's email sent it to every hotel, and every forward sent it further.

Set side by side, the two ways of sharing differ in every question the Ledger asks.

Question	The file on S:\REVENUE	The attachment
Where is it?	One path, on one server	A sent item, 40 inboxes, and wherever each recipient saved it
Who can open it?	Holders of NetWare trustee rights on S:\REVENUE, set by an administrator	Every recipient, and anyone they forward it to
Who can change it?	Holders of write rights; changes reach every reader	Each holder, on a copy that no other reader sees
How is it replaced?	Save over it; readers see the new version	Send a new email and hope the old copy is deleted
How is it removed?	Delete one file (and its backups)	Find every mailbox, folder and PC that kept a copy

Head office could not have chosen the left column in 1998. The hotels were not on the network, and a path on S: meant nothing to a PC in another city. A copy, sent by email, was the practical way to reach them. The cost of that choice was not the copy itself but the missing record that should have travelled with it: which file this was, which file it replaced, and who answered for it.

When the format breaks

Copies also break when the programs that read them differ. In 1998 the reservations manager saved the rate sheet in Excel 97. Several hotels still ran older versions of Excel, and they could not open the Excel 97 attachment in their version of Excel until they upgraded. In the meantime some asked head office for a copy in an older format.

Microsoft saw the problem coming. Its Office 97 resource kit chapter on upgrading to Excel 97 for Windows and Excel 98 for the Macintosh told administrators that "the Excel 97 file format differs from that of previous versions of Excel", and that a workgroup "upgrading gradually" would have users who needed to share workbooks with Excel 95 users. It offered two strategies: save in a dual format, "Microsoft Excel 97 & 5.0/95 Workbook", which held one data stream for each version, or save in the Excel 5.0/95 format. Each had a cost. In the dual format, an Excel 95 user who saved the file kept only the older stream, and "the 97 or 98 data stream is permanently lost". In the older format, the same kit lists what does not survive the trip back: Excel 97 allowed 65,536 rows per worksheet, and saved in the Excel 5.0 or 95 format, "data in rows below row 16,384 are truncated".

The rate sheet never came near 16,384 rows. The point is what the workaround did to the copies. A hotel that could not open the file asked head office for an older version, which became another copy with another name. A hotel that opened the dual format and saved it in Excel 95 now had a copy that no longer matched head office's. A hotel that upgraded in March and its neighbour that upgraded in June spent a season on different files. None of this showed in a folder listing. Each copy looked like a rate sheet.

WHAT HASN'T CHANGED

Microsoft's support page on file formats, read in November 2026, still lists the "Excel 97 - Excel 2003 Binary file format (BIFF8)" as a format Excel can save, and describes the .xlsx workbook as "the default XML-based file format for Excel 2010 and Excel 2007". The binary format of 1997 and the XML format of 2007 both still open. Each format change gives every organization a season in which two versions of the same file circulate, and a chance to ask which one is the record.

"final", "final v2", "final GM"

Once copies multiply, people use the only metadata they control: the file name. The 1987 sheet was RATES.WK1. The MS-DOS FAT file system allowed a name of at most eight characters and an extension of three (Chapter One). Microsoft documents that the Windows FAT and NTFS file systems "are not limited to 8.3 file names, because they have long file name support", and by 1998 the hotels' Windows PCs used it. A general manager who changed the festival weekend saved the season's file as RATES98 final GM.xls. Another, editing it again after a call from head office, could save "final v2". Across 40 hotels the names were a record of good intentions: each one says "this is the version I trust", in a form no system reads.

A file name is a claim by the last person who saved it. "Final" means final for that person on that day. It does not say final for whom, compared with what, or approved by whom. Two files named final can disagree, and a file named draft can be the one head office loaded. The name also travels badly: when a file is attached, saved and renamed, the word "final" stays in the name while the facts behind it change. Chapter Two calls this a locator, not an identity. The working rule here is simpler: a name can help a person find a file; it cannot tell anyone which copy is authoritative.

That leaves the person at the desk as the real reader of the era. A front office clerk at a hotel opens the rate sheet in the morning to quote a price. She does not know whether head office has sent a newer file, whether her general manager changed a rate last week, or whether the

copy on her PC is the one from the email or one a colleague saved over it. She reads the cell in front of her. The sheet has no field for "superseded on" and no link back to the master. The authority she trusts is the date of the email it came in, which she can no longer see.

The hotels also became readers of each other's habits. When a general manager moved to another hotel, she took her way of naming and keeping files with her, and sometimes her files. The harbor hotel's sales coordinator had linked a desktop database to the hotel's copy in 1996 (Chapter Five). In 1998 the sheet arrived by email under a new name, last season's file was cleared from the hotel PC, and the link broke. The file had moved from a location to a message, and every tool that pointed at the old location now pointed at nothing, or at an old copy.

Storage is cheap; reconciliation is not

The cost of a copy splits in two. Storing it costs almost nothing, and the cost kept falling through the era. Reconciling it costs people's time, every time two copies disagree. The rate sheet shows the split. Mail storage for 40 attachments a season was close to zero. Head office spent about three days each season (an illustrative figure for the thread) comparing returned copies, chasing hotels that had changed rates, and deciding which edits to accept.

Cost	Who pays	When it shows
Disk and mailbox space for each copy	IT, as a budget line	Rarely; small per copy
Reconciling copies that disagree	Head office staff, in hours	Every season
A price quoted from an old copy	The hotel and the guest	When a guest complains
Finding every copy for a legal or privacy duty	Legal, privacy, IT	Years later, under a deadline

The first line is the one budgets saw. The others were paid in staff time and in errors, and nobody added them up. That is why the copy won: its price was visible and small, and its cost was spread across many people and many years. The last row has no figure in 1998. It acquires one in Chapter Eleven, when an erasure request has to reach every mailbox attachment since 2004.

The habit also outlived the reason for it. The seasonal email continued until the revenue system replaced the workbook in 2022. By then the group had about 60 hotels, a cloud library and links that could have replaced the attachment years earlier, but the email was how the hotels expected to receive rates, so it kept going. When the workbook was retired,

some hotels kept emailing their leftover copies to new staff and to each other, because nothing told them to stop. Those leftovers are part of why the group found 23 copies of the rate sheet in 2026 (Chapter Twelve counts them). A distribution habit is a system with no owner: it runs until someone switches it off, and nobody had been given the switch.

OUR VIEW, AS OF NOVEMBER 2026

The cheapest fix in 1998 would have been two sentences in each email: "This file replaces the previous season's. The master is the reservations manager's copy at head office (still named RATES94.XLS on the server); make the local changes you need, and send them to her, because only her copy counts." It would not have stopped copies. It would have told every holder who owns the master and where changes go, which is most of what the Ledger asks for.

THE HOP · 2 · 1998 · HEAD OFFICE FILE SERVER TO EMAIL ATTACHMENT TO 40 HOTELS, ONE COPY PER HOTEL

Record	The rate sheet (hotel and retail group), hop 2, 1998
From	Head office file server
To	Email attachment to 40 hotels, one copy per hotel
Mechanism and schedule	Each season the sheet is emailed to every hotel; general managers save and edit their own copies; the practice continues until the 2022 replacement
Format change; what was lost or assumed	OWNERSHIP Excel 97 workbook (hotels on older Excel cannot open it until they upgrade); long file names such as RATES98 final GM.xls. **Each hotel saves and edits its own copy; head office's S: copy stays the master by habit, not by any record**
Accountable owner	None recorded for the hotel copies; the reservations manager for the S: copy
Who can change it	Every general manager, and anyone a copy is forwarded to
Classification; retention or erasure duty	Commercial pricing data; no retention rule
Read by (people, apps, models, agents)	40 general managers and their front office staff
Cost per month; basis	Per month: mail storage near zero; per season: about three days of reconciliation (basis: staff time) (illustrative)

Hop 2, in 1998, took the rate sheet from the head office file server to an email attachment sent to 40 hotels, one copy per hotel. Each season the sheet went out again, and the practice continued until the 2022 replacement, by which time the group had about 60 hotels. It was an Excel 97 workbook, so hotels on older Excel could not open it until they upgraded, and it carried long file names such as RATES98 final GM.xls. Each general manager saved and edited a copy. Head office's S: copy stayed the master by habit, not by any record. No owner was recorded for the hotel copies; every general manager could change one, and so could anyone a copy was forwarded to. The readers were 40 general managers and their front office staff. Mail storage cost almost nothing; reconciliation took about three days each season. The loss is OWNERSHIP: from this hop on, the rate sheet is a family of copies, and only one of them has an accountable owner.

ERA CARD

Reader. The person at the desk, now at a hotel as well as at head office, reading a copy that arrived by email and cannot show where it came from.

Security. Server rights protect the share; nothing protects the attachment once it is sent. Every recipient can forward, rename and keep it.

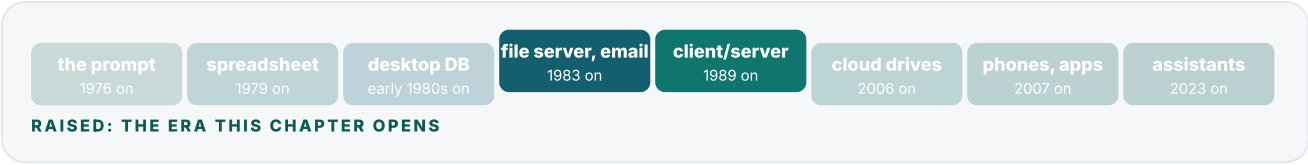
Cost. Storage near zero; reconciliation measured in days each season; the cost of finding every copy is deferred, not avoided.

AI. Old attachments and saved copies stay in mailboxes and folders that a workplace assistant can later read, each one looking like the rate sheet.

Monday rule: pick the five files your teams email most often, and for each one name the master location and its owner, and replace the attachment with a link to that location where your tools allow it.

CHAPTER SEVEN

CSV, the Format That Never Died



In 2003 the hotel group rented a booking engine, a web service that let guests reserve rooms online. The engine took rates through an extranet upload, and the upload took one format: a CSV file. Head office exported one CSV per hotel from the rate workbook. Each hotel's front office manager opened the file, adjusted it and uploaded it. Within a season, room type codes had lost their leading zeros, dates had swapped day and month, and two hotels were publishing a 189-euro room at 18900 until a guest complained. Nobody had made a mistake that the tools could see. Each program had read the same characters by its own rules. This chapter covers the plain text file of comma-separated values: what it is, why it outlived every format around it, and where its meaning leaks. The decision it changes: **no CSV enters a system of record without a written contract for its delimiter, decimal mark, date order, encoding and code columns**.

A format older than its specification

A CSV file is text. Each line is a record, and a separator, usually a comma, divides the line into fields. Almost every data tool can write one and read one, which is why spreadsheets, databases, booking engines and reporting tools all offer it. It needs no licence and no library. A person can open it in a text editor and see what is there.

For most of its life it had no written definition. RFC 4180, published as an RFC in October 2005, set one down, and said why it was needed: "there is no formal specification in existence, which allows for a wide variety of interpretations of CSV files." It is an Informational document, and it says so: "It does not specify an Internet standard of any kind." It describes the common practice: each record on its own line, ended by a line break (CRLF); "an optional header line" as the first line; fields that contain line breaks, double quotes or commas wrapped in double quotes. It notes that "common usage of CSV is US-ASCII", with other character sets allowed through a charset parameter.

The dates matter for the thread. The hotels' upload began in 2003, two years before RFC 4180 existed, so it could not have followed it. The engine's own upload guide was the only contract, and the hotels' spreadsheets did not read it. Even after 2005, an Informational RFC binds nobody: it records a common shape, and programs remain free to differ.

Ten years later the World Wide Web Consortium went further. Its *Model for Tabular Data and Metadata on the Web*, a W3C Recommendation dated 17 December 2015, makes the same observation in its section on best practice CSV: "There is no standard for CSV, and there are many variants of CSV used on the web today." Its answer is metadata kept beside the file: columns "may be annotated with a datatype", and a column can be marked "required", meaning its cells must not be empty. The model is sound. The hotels' desk tools wrote no such metadata. In practice a CSV still arrives as text with a file name, and the reader supplies the rest.

What a spreadsheet writes is also less than what it holds. Microsoft's page on supported file formats says the CSV format "saves only the active sheet". Since the 1994 rebuild the rate workbook had a tab per hotel and, in its notes, the conditions of each rate. The CSV for a hotel carried one table of codes and numbers. Whatever lived on other sheets, in comments or in formatting stayed behind. The rate conditions had sat outside the sheet since 1987 (the CONTEXT loss of hop 0), and the CSV gave them no way back in.

Delimiters, decimals and the regional setting

The comma in "comma-separated" is a convention, and it collides with another. In many European countries the comma is the decimal mark: 189,00 euros means one hundred and eighty-nine euros. A file that used commas for both jobs could not be read. The desk tools resolved it with the operating system's regional settings.

Microsoft's support page on importing and exporting text files explains the mechanism for Excel. "When you save a workbook as a .csv file, the default list separator (delimiter) is a comma." The separator can be changed in Windows' region settings, with a warning that it "will cause a global change on your computer, affecting all applications". And changing the decimal separator to a comma "forces Excel to use a semi-colon for the list separator". When Excel opens a CSV, it "uses the current default data format settings to interpret how to import each column". The file does not say which settings wrote it. The reader's PC decides.

So the same rate row exists in two dialects. Take room type 01 on 4 May 2003 at 189 euros. A PC set for the United States writes `01,5/4/2003,189.00`: commas between fields, month first, a decimal point. A PC whose regional settings use day-first dates and a decimal comma writes `01;04/05/2003;189,00`. Both are correct by their own rules. Whatever head office sent, an EU hotel that imported the file and saved it again wrote it in the hotel's dialect: semicolon-delimited, day-first dates, decimal commas. The engine parsed with US settings.

It split the fields on the semicolon correctly, then read each field as a US number and a US date. In a number it treated the comma as a thousands separator and dropped it, so 189,00 became 18900. That was this engine's rule, not every parser's: another might have rejected the value or kept it as text, which is why the rule belongs in a contract. Most hotels checked their published prices after the first upload and caught the hundredfold rates. Two did not, and published 18900 until a guest complained.

The defence is not a better spreadsheet. It is to state the dialect in the contract (separator, decimal mark, thousands mark, quoting) and to check every file against it before the upload, rejecting a file that does not match rather than guessing. The engine guessed.

The 132 dates that swap silently

A date written as 04/05/2003 is the fourth of May to a reader who expects day first, and the fifth of April to a reader who expects month first. The text cannot tell them apart. A date such as 25/05/2003 can only be day first. A strict US parser rejects it, and the error is loud. This engine was not strict: when a date could not be read month first, it read it day first. Those dates landed correctly, and nothing warned anyone about the rest. The dangerous dates are the ones that parse both ways.

Count them once. A date is ambiguous when both the day and the month are 12 or less. There are 12 x 12 = 144 such day and month pairs in a year. In 12 of them the day equals the month (01/01, 02/02 and so on to 12/12), and those read the same either way. That leaves 144 minus 12 = 132 dates a year that a day-first writer and a month-first reader will swap without any error. For a rate sheet with a rate for every day, that is up to 132 days of rates a year published on the wrong dates, each one a valid date, none of them flagged.

The swap also travels. The engine now holds, against 5 April, the rate the hotel set for a stay on 4 May. The record looks correct and means something else, and any copy taken from the engine carries it. Nothing downstream can repair it without going back to the hotel's intent. The fix belongs at the boundary: write dates in one stated order, with a four-digit year first (2003-05-04), and refuse any date field whose order is not declared.

Zeros, long numbers and precision

The hotels' room types ran from 01 to 12. In the hotel's sheet they were codes, typed as text. In the CSV they were characters: a zero and a one. The trouble came when a front office manager double-clicked the CSV to open it in a spreadsheet, adjusted a rate and saved it again. (A manager who opened it through the text import wizard and marked the column as text kept the zeros.) Microsoft's support page states the behaviour plainly: "Excel

automatically removes leading zeros, and converts large numbers to scientific notation". Room type 01 became 1. Room types 01 to 09 all lost their zero; 10, 11 and 12 did not. The engine then received codes that no longer matched its room table.

The same page gives the limit for long numbers: "Excel has a maximum precision of 15 significant digits", and in a longer number "any numbers past the 15th digit are rounded down to zero". A 16-digit identifier read as a number is shown as something like 1.23457E+15 and stored with its last digit replaced. Microsoft's fixes are to format the column as text before entering data, or to use Get & Transform (Power Query) "to format individual columns as text when you import data". Formatting afterwards does not help. Microsoft now documents a set of Automatic Data Conversions options, read in November 2026, that can turn off "removing leading zeros from numerical text and converting to a number", truncating "to 15 digits of precision", and "converting a continuous string of letters and numbers to a date". The options change what Excel does on that PC. They do not change what the file says, or what the next person's Excel does.

The health system shows the cost outside hotels (an illustrative case). A clinic exports its waiting list as a CSV, with a medical record number per patient, assigned as seven digits with leading zeros: 0012345. A coordinator opens the file to sort it and saves it. The record number is now 12345. Uploaded to another system, it no longer matches the patient it named, and in a system where 12345 belongs to another patient, it matches someone else. A record number is an identity, not a quantity. It should not pass through a tool that treats it as one.

Precision broke a third way at the engine. Its rate field held whole currency units. In the files that carried a decimal point, from hotels whose regional settings used one, a rate of 189.50 was stored as 190. (In a decimal-comma file the same rate would have become 18950, the error above.) No program reported it. The rate sheet's cents existed in the workbook and nowhere after the upload.

FIGURE 7.1

One rate, several readers: each program reads the text by its own rules, and the meaning changes with them

One rate row (an EU hotel, room type 01, 4 May 2003, 189 euros) and one record number

Source head office workbook	`room 01 | 4 May 2003 | 189.00`	Room type a text code; a date value; a rate with cents
Saved by an EU hotel column imported as text	`01;04/05/2003;189,00`	Semicolons, day first, decimal comma; the file does not say so
Parsed by the engine US settings	`room 01 | 5 Apr 2003 | 18900`	Month first: 4 May becomes 5 April; comma dropped as a thousands separator
Another hotel's file double-clicked, resaved	`1;04/05/2003;189,00`	Leading zero removed: 01 becomes 1, a code the engine does not know
A different rate, 189.50 decimal-point file, whole units	`189.50 → 190`	Rounded on upload; no warning; the cents exist only in the workbook
Clinic export record number	`MRN 0012345`	Seven characters: an identity, not a quantity (illustrative)
Clinic export resaved in a spreadsheet	`MRN 12345`	Zeros removed; no longer matches the patient it named

An EU hotel's rate for room type 01 on 4 May 2003, at 189 euros: in the head office workbook, as saved under EU settings and as parsed by the engine's US settings; beside it, two other paths (another hotel's double-clicked and resaved file, and a decimal-point file in the engine's whole-unit field); below, a clinic's record number through the same reopen and save. Red rows are wrong readings.

Author's diagram based on Microsoft Support, "Import or export text (.txt or .csv) files" and "Keeping leading zeros and large numbers" (read 20 November 2026); the hotel row and the clinic record number are illustrative.

TYPE CONTRACT

Type, precision and rounding. Room type: text code of two characters, 01 to 12, not a number. Rate: decimal with two places; the engine field must hold cents, or the upload is refused.

Encoding and serialization. UTF-8 with a stated byte order mark policy; one delimiter stated in the contract; decimal point, no thousands separator, written by a fixed export template rather than by each PC's regional settings; fields quoted when they contain the delimiter.

Unit or currency. Euros for EU hotels, stated in a currency column, not implied by the hotel.

Time basis. Stay date in year, month, day order (2003-05-04), local to the hotel; no two-digit years.

Identity scope. Hotel code plus room type plus stay date plus rate code identifies a row; room type 01 is unique only within one hotel.

Absence. An empty rate means not on sale, and the contract says so. A rate of zero is refused, because the engine would publish it as a free room.

Authority. Head office's workbook is the source; a hotel's edited CSV is a working copy until head office accepts it.

Types, units and time are covered in depth later in this series. The point here is narrower: the CSV carried none of these seven answers, so every reader made them up.

Encodings and formulas

Text needs an encoding, the rule that turns characters into bytes. RFC 4180 noted that common usage was US-ASCII, which has no accented letters. A guest name such as Müller or a city such as São Paulo needs more, and UTF-8, a Unicode encoding, can carry them. But a CSV does not say which encoding it uses. Microsoft's guidance for Excel: "You can open a CSV file encoded with UTF-8 normally if it was saved with BOM (Byte Order Mark)." The byte order mark is a short sequence of bytes at the start of the file that signals UTF-8. Without it, Microsoft points users to Power Query or the text import wizard. Whether to write the mark is therefore a contract term, agreed with the reader, not a default.

A CSV can also carry instructions. The OWASP Foundation describes CSV injection, "also known as Formula Injection", which "occurs when websites embed untrusted input inside CSV files". A field that begins with an equals sign, a plus or minus sign, an at sign, a tab, a carriage return or a line feed can be read by a spreadsheet as a formula when a person opens the file. A guest's name or a comment typed into a web form can end up as a formula on an analyst's screen. OWASP's mitigations include wrapping each field in double quotes, prefixing a risky cell with a single quote (or, for Excel, a tab character inside the quoted field) and escaping double quotes, and it warns that Excel may remove quotes or escape characters when a file is saved and opened again, so that escaped formulas can become active. The control belongs with whoever writes the file: treat every field from outside as data, neutralize the leading characters before export, and test the result in the spreadsheet that will open it.

WHAT HASN'T CHANGED

Microsoft's import page, read in November 2026, still advises renaming a file "from .csv to .txt before you open it" to get the text import wizard, so that a person can choose each column's format, keep leading zeros or set a date order. The safe path through a CSV in 2026 is the one of 2003: tell the reader what each column is, before it guesses.

The four losses in the hotels' uploads share a cause, and the table shows it: each time, the file left a question open and a different program answered it.

What the CSV did not say	Who answered	What happened
Room type is a code, not a number	The spreadsheet that reopened the file	01 to 09 became 1 to 9
Dates are day first	The engine's US settings	Up to 132 dates a year swapped
The comma is a decimal mark	The engine's US settings	189,00 published as 18900
Rates carry cents	The engine's rate field	189.50 stored as 190

Custody made the losses hard to trace. Each hotel's front office manager edited and uploaded that hotel's file, and the engine vendor's support staff could also change rates on request. Head office kept the workbook but not the files that were uploaded from it. When the mispriced dates appeared, nobody could say which file, from which PC and with which settings, had produced them, because the only copy that mattered was the one the engine had already parsed. A CSV upload is a hop between systems. It deserves what any hop gets in the Ledger: a kept copy of what was sent, the settings it was read with, and a named person who sent it.

THE HOP · 3 · 2003 · HEAD OFFICE WORKBOOK TO RENTED BOOKING ENGINE (EXTRANET UPLOAD), ONE CSV PER HOTEL

Record	The rate sheet (hotel and retail group), hop 3, 2003
From	Head office workbook
To	Rented booking engine (extranet upload), one CSV per hotel
Mechanism and schedule	Head office exports one CSV per hotel; each hotel's front office manager edits it and uploads it to the booking engine's extranet; web rate codes are created (the ancestors of today's BAR codes)
Format change; what was lost or assumed	IDENTITY TIME MEANING PRECISION CSV text; EU hotels save semicolon-delimited files with day-first dates and decimal commas under their regional settings; the engine parses with US settings. **Room types 01 to 09 lose their leading zeros when a hotel opens and resaves its CSV in a spreadsheet; 04/05 is read as 5 April, so up to 132 dates a year swap silently; in the semicolon-delimited files the comma in 189,00 is read as a thousands separator, so two hotels publish 18900 until a guest complains; the engine's rate field holds whole units, so 189.50 becomes 190. The engine reads dates month first and falls back to day first when the month would exceed 12, so only ambiguous dates swap; most hotels caught the 18900 rates on checking their prices, two did not; 189.50 comes from files written with a decimal point**
Accountable owner	Head of distribution (engine contract); front office managers (uploads)
Who can change it	Each hotel's front office manager; the engine vendor's support staff
Classification; retention or erasure duty	Commercial pricing data published to the web
Read by (people, apps, models, agents)	Guests on the web, the booking engine, channel partners
Cost per month; basis	Per month: the engine fee per booking; one-off: rework of the mispriced dates (not monthly) (illustrative)

Hop 3, in 2003, is a side branch: from the head office workbook to the rented booking engine, one CSV per hotel. Head office exported each file; each hotel's front office manager edited it and uploaded it to the engine's extranet, and web rate codes were created, the ancestors of today's best available rate (BAR) codes. The EU hotels saved semicolon-delimited files with day-first dates and decimal commas under their regional settings, and the engine parsed with US settings. Four losses happened at once. IDENTITY: room types 01

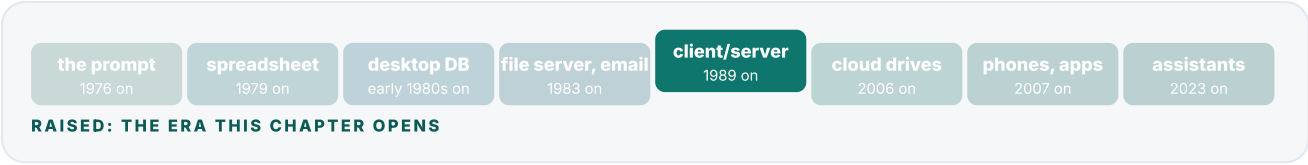
to 09 lost their leading zeros when a hotel opened and resaved its CSV in a spreadsheet. TIME: 04/05 was read as 5 April, so up to 132 dates a year swapped silently. MEANING: in the semicolon-delimited files the comma in 189,00 was read as a thousands separator, and two hotels published 18900 until a guest complained. PRECISION: the engine's rate field held whole units, so 189.50 became 190. The head of distribution owned the engine contract; the front office managers owned the uploads; the engine vendor's support staff could also change rates. The readers were guests on the web, the engine and channel partners.

> **ERA CARD**
>
> **Reader.** A program at the other end of an upload, reading text by its own regional rules, and the front office manager who opened the file in between.
>
> **Security.** An extranet login per hotel; no check of what the file contained; fields from outside could carry formulas.
>
> **Cost.** An engine fee per booking; the rework of mispriced dates; and the hours spent finding which file, from which PC, caused it.
>
> **AI.** An assistant that retrieves a CSV can inherit every guess the last spreadsheet made: a code without its zero, a date in the wrong order, a rate without its cents.

Monday rule: for every CSV your systems accept, publish a one-page contract (delimiter, decimal mark, date order, encoding, text columns) and reject files that break it at the door instead of repairing them inside.

CHAPTER EIGHT

Client/Server and the Pipe to the Desk



In 2004 an analyst at the hotel group's head office spent an afternoon making the rate sheet smarter. The group's customer relationship management system, the CRM, ran on its own SQL Server database, and it held the corporate accounts, their negotiated rates and the names of the people who signed them. The analyst installed an ODBC driver, wrote a query, and pointed a new tab in the rate workbook at the database. The tab, called CORP and placed as the workbook's first sheet, held a small table of corporate rates and client contacts, and it refreshed each time the workbook opened. On a second tab, VIP, the analyst pasted the group's top loyalty guests, copied from the same CRM. It was good work and it saved hours. It also moved data that only the sales team could read in the CRM into a file that every hotel received each season. This chapter covers client/server computing and the pipe it built from the database to the desk. The decision it changes: **every live connection from a system of record into a desk file gets an owner, a purpose, an audience and an end date, and the copy says so when the pipe stops**.

The server comes to the network

Client/server split the work of a database in two. A server program held the data and answered queries; client programs on desk PCs sent the queries and showed the answers. Before it, a desk user either typed data into a local file or asked the department that ran the central system for a report. After it, a PC on the network could ask the database directly.

Microsoft's history records the moment for its own products: on 3 May 1989, "Microsoft and Ashton-Tate announce the shipment of the Microsoft SQL Server 1.0", the result of "a joint development effort of Ashton-Tate, Microsoft, and Sybase", described as "a powerful, relational database server for PC-based Local Area Networks (LANs)". The phrase that matters is the last one. The relational database, with its schema, keys and transactions, now sat on the same network as the spreadsheet. The two were a cable apart.

The hotel group's CRM was one such system: an on-premises application on SQL Server, owned by the sales team. Inside it, a corporate client's contact details were readable by the sales team. That rule lived in the CRM's permissions. It was the authorization boundary for the data, and it was about to be crossed by a query.

ODBC: one interface, many drivers

A desk tool that wanted to query a database needed to speak that database's language. Open Database Connectivity removed the need. Microsoft's history lists the "Open Database Connectivity (ODBC) Software Development Kit 1.0" among its 1992 releases. Its current documentation defines ODBC as "a specification for a database API", independent of any one database or operating system, and "based on the CLI specifications from Open Group and ISO/IEC". Database vendors write drivers; applications call the functions in those drivers; "a Driver Manager manages communication between applications and drivers". To the end user, the same page notes dryly, ODBC is "an icon in the Microsoft Windows Control Panel".

That icon is where a data source was set up: a name, a driver, a server, a database and, often, a login. Once it existed, Excel and Access could use it. Excel could place the result of a query in a range of cells and refresh it on demand. Access could link a table on the server as if it were local (Chapter Five covers linked tables and what happens when their target moves). The same documentation adds a caution worth repeating: ODBC "is designed to expose database capabilities, not supplement them". The database applies its permissions while a query runs. Once the rows land in a workbook, the rules around them stay behind: who may see a row, why it was collected, how long it may be kept.

Refresh on open and the credentials inside

Two settings decided how the CORP tab behaved, and both are still in Excel. Microsoft's page on connection properties, read in November 2026, describes the first: "Select this check box to automatically refresh external data when you open the workbook." With it set, each opening reran the query, as long as the PC could reach the server and the login worked. If the refresh did not run, the tab showed whatever the last refresh had left. A related option saves "the workbook with the query definition but without the external data"; the CORP tab did not use it, so the data travelled with every copy. Refresh ran only on the head office network. The hotels were not on it, so at a hotel each opening's refresh failed and the tab showed the rows as of the last save at head office, from the first season on.

The second setting is the login. The same page describes a "Save password" option that stores the user name and password in the connection file, and states: "The saved password is not encrypted." Its security note: "Avoid saving logon information when connecting to data

sources." The information "may be stored as plain text, and a malicious user could access the information to compromise the security of the data source." A workbook that refreshes for anyone who opens it asks each reader for a login, relies on each reader's own Windows login, or carries a saved one. A carried login turns every copy of the workbook into a key to the database, readable by whoever ends up with the file. Whichever way the CORP query logged in, the rows already saved in the file needed no login at all: any copy showed them to anyone who opened it.

Connection files have one more behaviour that matters later. Microsoft documents that when a workbook is set to use a connection file and that file "is not available, Excel resorts to the connection information that is saved in the workbook". The workbook keeps its own memory of where its data came from. When the source moves, the memory does not.

The extract as self-service

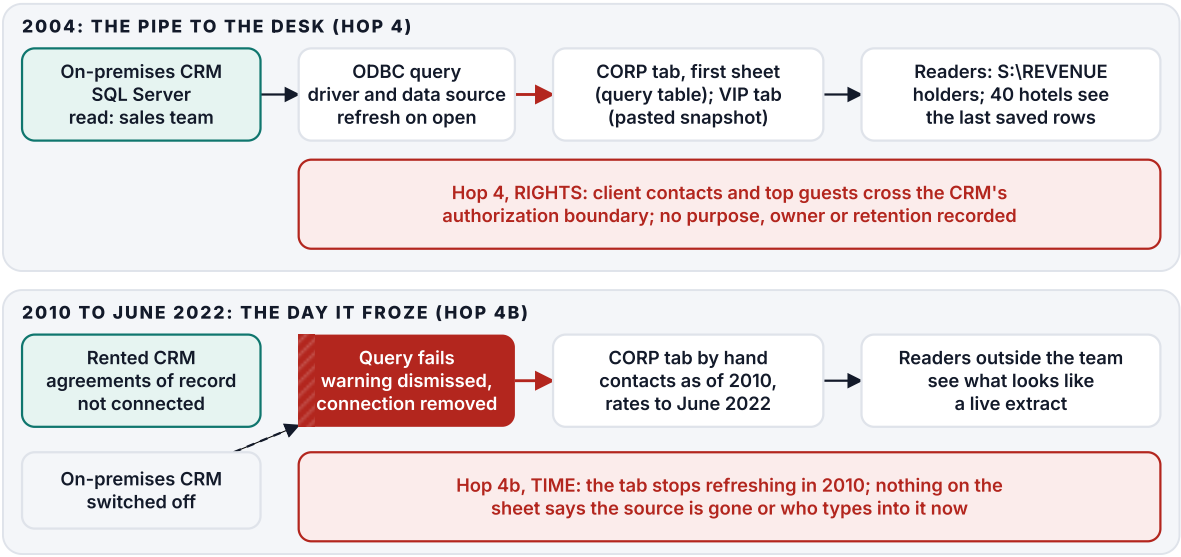
The CORP tab was an extract: a copy of rows from a system of record, taken by the person who needed them, without a request to the system's owner. That is what made it useful. The analyst did not need a project, a budget or a change ticket. The cost was an ODBC driver and an afternoon.

What the extract left behind was the boundary. In the CRM, a client's contact was visible to the sales team. In the workbook, it was visible to everyone with rights to S:\REVENUE, and from the next seasonal email (Chapter Six), to every hotel that received the sheet and anyone a hotel forwarded it to. No purpose was recorded for the copy, no owner for the tabs and no retention period. The VIP tab was worse in one way: it was a pasted snapshot of guests' names and preferences, with no connection back to anything, so even the CRM could not tell that the copy existed.

The law was not silent. The group's EU hotels were then subject to the national laws that implemented Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, the instrument the General Data Protection Regulation later replaced. This chapter does not argue how a regulator would have judged the tab. It notes what the group could not have shown for it: a recorded purpose, an owner, or a reason for 40 hotels to hold client contacts. The failure was not the driver or the spreadsheet. It was the absence of any rule that said a copy of CRM data needed the same answers as the CRM.

FIGURE 8.1

The pipe to the desk and the day it froze: a query carried CRM data past its boundary in 2004, and kept looking live after 2010



Hop 4 (2004): an ODBC query moves sales-only CRM data into the CORP tab that every hotel receives; hop 4b (2010): the source is switched off, the connection later removed, and the tab is kept by hand (contacts as of 2010, rates to June 2022) while it still looks like an extract.

Author's diagram based on the book's running example (hops 4 and 4b); the hotel group and its systems are illustrative. ODBC roles from Microsoft Learn, "What Is ODBC?"; refresh and password options from Microsoft Support, "Connection properties" (read 20 November 2026).

When the pipe breaks, the copy still looks live

In 2010 two changes met. The workbook was saved as Rates master.xlsx, in the Office Open XML format that was new to the group; Microsoft describes .xlsx as "the default XML-based file format for Excel 2010 and Excel 2007". In the same year the CRM moved to a rented service, and the on-premises database was switched off. The CORP query still pointed at it.

The next refresh failed. Head office staff dismissed the warning, as hotel staff had dismissed theirs every season since 2004. Later someone removed the connection, so the warning stopped. The tab kept its last rows, and the revenue team maintained it by hand from then on, typing in rates while the client contacts stayed as they were in 2010, until the last save in June 2022. To the revenue team it was a hand-kept list. To everyone else it looked exactly as it had in 2004: a tidy table with a header, client names, rates and contacts, the shape of a database extract. Nothing on the sheet said that the source was gone, when the last refresh had run, or who now typed into it. The VIP tab, a pasted snapshot from the start, went on as it was: guest names and preferences in every copy.

That is the particular risk of a broken pipe. A file that was hand-made from the start carries its own warning, because readers know a person made it. A file that was once a live extract keeps the authority of the system it came from after the system has left. The corporate agreements of record had moved to the rented CRM, which was not connected to the workbook and, later, was not connected to the workplace assistant either. Chapter Thirteen shows what that assistant made of the tab.

TYPE CONTRACT

Type, precision and rounding. Negotiated rate: decimal with two places, typed by hand since 2010; no check against the agreement.

Encoding and serialization. A plain range in an .xlsx sheet; the query table converted to a range and its connection removed; nothing on the sheet shows it was once a query.

Unit or currency. Euros for the EU hotels, implied by the hotel; not stated per row.

Time basis. Last refreshed from the CRM in 2010, so contacts as of 2010; rates typed by hand to June 2022; the sheet shows neither date.

Identity scope. Client name as typed; no CRM account key, so a row cannot be matched to the agreement of record.

Absence. Conditions such as the festival blackout are absent, and nothing marks them as missing.

Authority. None since 2010: the rented CRM holds the agreements; the tab is a working copy that looks like a reference copy.

The tools that replaced the 2004 query have improved, and they are a pointer here, not a chapter. Microsoft describes Power Query as "a data transformation and data preparation engine" that saves the steps "as a repeatable query that you can refresh", available in Excel and Power BI, and it notes that queries can be refreshed manually, "by using scheduled refresh capabilities in specific products such as Power BI, or programmatically". Scheduled refresh adds what the 2004 tab lacked: a service that runs the query and reports when it fails. Microsoft documents, as of November 2026, that Power BI "deactivates your refresh schedule after four consecutive failures" or when it detects an error such as expired credentials, and that a schedule paused after two months of inactivity emails the semantic model's owner. A failed pipe now has a status and an owner who hears about it.

The pipe still does not carry the purpose, the audience or the end date. A refreshed report that lands in a file can still be copied, emailed and saved after its source is retired, and the copy does not refresh. Chapter Nine follows the workbook into the cloud library, where the CORP tab acquired a much larger audience.

THE HOP · 4 · 2004 · ON-PREMISES CRM ON SQL SERVER TO THE CORP TAB, THE WORKBOOK'S FIRST SHEET (ODBC QUERY, REFRESHED ON OPEN), AND A VIP TAB (PASTED SNAPSHOT)

Record	The rate sheet (hotel and retail group), hop 4, 2004
From	On-premises CRM on SQL Server
To	The CORP tab, the workbook's first sheet (ODBC query, refreshed on open), and a VIP tab (pasted snapshot)
Mechanism and schedule	An analyst adds an ODBC query that pulls corporate accounts, negotiated rates and contacts into a CORP tab, and pastes the top loyalty guests into a VIP tab
Format change; what was lost or assumed	RIGHTS Excel query table over ODBC; contact names, business emails and phone numbers; guest names and preferences. **Client contacts and top guests, readable in the CRM only by the sales team, become readable by everyone with rights to S:\REVENUE and by every hotel that receives the sheet; no recorded purpose, owner or retention period (the EU hotels then under national laws implementing Directive 95/46/EC)**
Accountable owner	None recorded for the tabs
Who can change it	The analyst who wrote the query; anyone who can save the workbook
Classification; retention or erasure duty	Personal data of business contacts and guests; commercial terms
Read by (people, apps, models, agents)	Head office, 40 hotels, anyone a copy reaches (the hotels read the rows as of the last save at head office; refresh ran only on the head office network; at the hotels each opening showed a failed refresh)
Cost per month; basis	One-off: an ODBC driver and about 4 hours of work, an afternoon (not monthly) (illustrative)

Hop 4, in 2004, ran from the on-premises CRM on SQL Server to two workbook tabs: CORP, an Excel query table over ODBC refreshed on open, and VIP, a pasted snapshot. An analyst added the query, which pulled corporate accounts, negotiated rates and contacts (names, business emails and phone numbers), and pasted the top loyalty guests with their names and

preferences. The loss is RIGHTS, and it is an authorization boundary crossed: client contacts and top guests, readable in the CRM only by the sales team, became readable by everyone with rights to S:\REVENUE and by every hotel that received the sheet, with no recorded purpose, owner or retention period, while the EU hotels were under national laws implementing Directive 95/46/EC. No owner was recorded for the tabs; the analyst who wrote the query, and anyone who could save the workbook, could change them. Readers: head office, 40 hotels and anyone a copy reached. Cost: an ODBC driver and an afternoon of work.

THE HOP · 4B · 2010 · RATES94.XLS AND ITS DESCENDANTS TO RATES MASTER.XLSX (EXCEL 2010)

Record	The rate sheet (hotel and retail group), hop 4b, 2010
From	RATES94.XLS and its descendants
To	Rates master.xlsx (Excel 2010)
Mechanism and schedule	The workbook is saved in the new format; the CRM moves to a rented service and the on-premises database is switched off
Format change; what was lost or assumed	TIME Office Open XML workbook (.xlsx); the CORP query's ODBC source no longer exists. **The CORP query fails on refresh; staff dismiss the warning, then remove the connection and maintain the tab by hand (contacts frozen at 2010; rates typed by hand) until the last save in June 2022; to readers outside the revenue team it still looks like a live extract. Corporate agreements of record move to the rented CRM, which is never connected to the workbook or, later, to the assistant**
Accountable owner	Head of revenue (a new role, 2009)
Who can change it	The revenue team; anyone with rights to the share
Classification; retention or erasure duty	Personal data and commercial terms; no retention rule
Read by (people, apps, models, agents)	The revenue team; hotels receiving the seasonal email; anyone with rights to the share
Cost per month; basis	Per month: a few hours of manual upkeep (basis: staff time) (illustrative)

Hop 4b, in 2010, took RATES94.XLS and its descendants to Rates master.xlsx in Excel 2010, saved in the Office Open XML format, while the CRM moved to a rented service and the on-premises database was switched off. The CORP query's ODBC source no longer existed. The

format change did not cause the loss; the switched-off source did, and the Ledger keeps both in one row because they reached the same file in the same year. The loss is TIME: the query failed on refresh; staff dismissed the warning, then removed the connection and maintained the tab by hand (contacts frozen at 2010; rates typed by hand) until the last save in June 2022, and to readers outside the revenue team it still looked like a live extract. Corporate agreements of record moved to the rented CRM, which was not connected to the workbook or, later, to the assistant. The head of revenue, a new role from 2009, owned the workbook; the revenue team and anyone with rights to the share could change it; the revenue team and the hotels receiving the seasonal email read it. Upkeep cost a few hours a month.

ERA CARD

Reader. The desk user who queries the database directly, and every later reader of the file who cannot tell a live extract from a frozen one.

Security. The database checks the login at refresh; the copy has no checks at all. A saved password is stored unencrypted with the connection details and can travel with them.

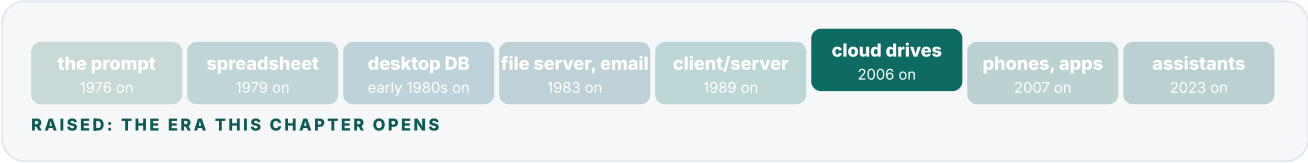
Cost. A driver and an afternoon to build; nothing to run; years of hand upkeep and exposure once the source moves.

AI. An assistant that retrieves the file can present the last rows of a dead query as current, because the file gives no sign that the pipe has stopped.

Monday rule: list the workbooks that hold a connection to a system of record, or held one, and for each record the source, the owner and the date of the last successful refresh in the file itself, removing saved passwords as you go.

CHAPTER NINE

From Shared Drive to Cloud Drive



In 2016 the hotel group moved its head office file server to the cloud. A third-party migration tool copied the folders into SharePoint Online and translated each folder's groups into permissions on the new libraries. Near the end, the tool reported a list of groups it could not resolve: old groups, nested groups, groups whose members had left. The plan had a date, and the team met it by mapping every unresolved group to one custom group, All staff (internal). The Revenue library, with the rate workbook in it, became readable by about 4,000 hotel and head office staff. Nothing happened to the workbook that weekend. What changed was the rule: from then on, membership of one group decided who could read it, and that group could grow without anyone changing the library. This chapter follows the move from the shared drive to the cloud drive, the arrival of the sharing link, and the change it made to the old question. On a file server you asked which copy was right. In a cloud library you must also ask who can open the one file, through which grant. The decision it changes: **a migration's permission mapping is a recorded decision with an owner, checked against the people it grants, not only against the folders it moves**.

The library, the personal drive and the sync client

Three products carried the desk file into the cloud era, each with its own history. Microsoft announced the release to manufacturing of SharePoint Portal Server on 28 March 2001, describing it as a portal that "combines enterprise-level search functionality developed by Microsoft research, with integrated document management tools". That server ran in a company's own data centre, like the file server it would slowly replace. A decade later, on 28 June 2011, Microsoft launched Office 365, which brought together Office, SharePoint Online, Exchange Online and Lync Online as a hosted service. The library the group migrated into in 2016 is that hosted descendant.

The personal drive for work arrived by a separate road. In January 2014 Microsoft announced that its consumer storage service, SkyDrive, would be renamed OneDrive, and that SkyDrive Pro, the business service, would move to "the new OneDrive for Business". At its SharePoint Conference on 3 March 2014, Microsoft said businesses could buy "OneDrive for Business (formerly SkyDrive Pro)" as a stand-alone product from April, and that SharePoint Server 2013 customers could turn it on in the cloud. The rename of SkyDrive was the consumer product's story; the business drive grew out of SharePoint. Google announced Google Drive on 24 April 2012, and Google Workspace customers run the same pattern of shared drives, personal drives and links.

Alongside the sanctioned services ran file sync services such as Dropbox and Box, installed by teams that wanted a folder they could share with a partner by Friday. Where nobody approved them, they were shadow IT: copies in a store the organization did not configure, could not search and might not know about. This book does not judge those products. It counts their copies, which belong in the copy census of Chapter Twelve.

What changed from the file server? Three things matter for records. First, the library keeps a **version history**. Microsoft documents, as of November 2026, that versioning is on by default for a new library or list in Microsoft 365 and "will automatically save the last 500 versions of a document"; administrators can choose an Automatic setting that keeps recent versions and thins older ones while staying within that 500 limit, or set their own count and expiry. Version history is useful for undoing a bad edit. It is not a backup, and it is a copy store of its own: every saved version of the rate workbook since the 2016 migration, with the rows deleted between versions, sits there until the limit trims it. Chapter Eleven returns to that when an erasure request reaches only the current file. Second, the library has **inherited permissions**: a file takes the library's access unless someone gives it narrower or wider access of its own. Third, every file has a **link**. Identity, the directory and the management plane behind these grants are the subject of *From Silicon to Screen*, Book 3 of *Every Pixel Matters*; this chapter only follows what the grants did to one workbook.

Microsoft Teams joined the same storage rather than adding a new one. Microsoft documents that files uploaded to a channel are stored in the team's SharePoint site, and that "files you send in chat are stored in your OneDrive for Business and only shared with the people in that conversation". A spreadsheet uploaded to a chat from a PC is therefore a file in the sender's personal drive, shared by a link with the chat's members; the sender's drive now holds a copy that the library's owner did not make and may never see. A file shared from a library travels as a link instead.

Three audiences for a link, and their defaults

A sharing link answers one question: who can open this item if they receive the link. Microsoft's documentation describes three audiences that widen access (a fourth setting, described below, limits a link to people who already have access and so grants nothing new), and the organization's settings decide which of them users may choose and which one the sharing dialog offers first. Microsoft documents the behaviour summarized in Table 9.1, as of its pages dated 30 June 2026 and 6 November 2026.

Link audience	Who can open it	Forwarding	Notes
Specific people	Named recipients, inside or outside the organization, who must authenticate as the person named	Forwarding grants nothing to anyone else	Microsoft calls it the most restrictive option; it "impedes broad internal sharing"
People in your organization	Anyone who signs in as a member of the directory	Works for any member it reaches; not for guests	Access is established when a recipient redeems the link; Teams and Outlook may redeem it automatically
Anyone with the link	Anyone who has the link, without signing in	Works inside and outside; anonymous use "can't be audited to a named user"	Available only when the organization's external sharing is set to Anyone

Two layers set the defaults. At the organization level, an administrator picks which of the three types the sharing dialog shows by default for files and folders, and each site can be the same or more restrictive. In November 2026 Microsoft also describes a single "hero link" per item: one URL whose audience and permission the user can change later "without needing to create a new URL". Its usual starting scope is "Only people added to the file have access", which grants nothing beyond the access a person already has; when a user first opens access through the hero link, "Can view" is the default permission, and Microsoft states that administrators cannot configure a different default for it. Narrowing the scope "applies to links already sent". Older links keep working and appear under other links.

Expiry is a setting, not a default. Microsoft documents that an administrator who allows Anyone links "can require all Anyone links to expire, and specify the maximum number of days allowed", and can limit them to view only; if the maximum is shortened, existing links update to the shorter period. Users can also create extra links with an expiration date or a

password. Where no administrator required expiry and no user set one, an Anyone link stays valid until someone removes it. The general manager's 2019 link, later in this chapter, was of that kind.

Two clarifications keep the risk in proportion. Microsoft documents that choosing Anyone "doesn't make the content globally searchable or accessible through Copilot, or grant access to everyone": a person must receive and redeem the link first. And an organization link does not grant organization-wide access either, until a recipient uses it. The broad internal grant that really does reach everyone is a group. Microsoft describes "Everyone except external users" as "a built-in SharePoint group that automatically includes all internal users but excludes any external guests", and offers data access governance reports, for organizations with the required licences, that list sites and items shared with it or with "Everyone". A custom group with the same reach does not carry that name, so a review that looks only for the built-in group passes it. That happened to the group's assistant pilot in 2025, as Chapter Thirteen tells.

Microsoft also states the rule that makes all of this one calculation: "A user's effective access is the highest permission granted by any source: direct access, a group, the parent folder or site, the hero link, or another sharing link." Who can open a file is the union of every path to it, and each person gets the highest level any path grants. Reducing access means finding and closing each path.

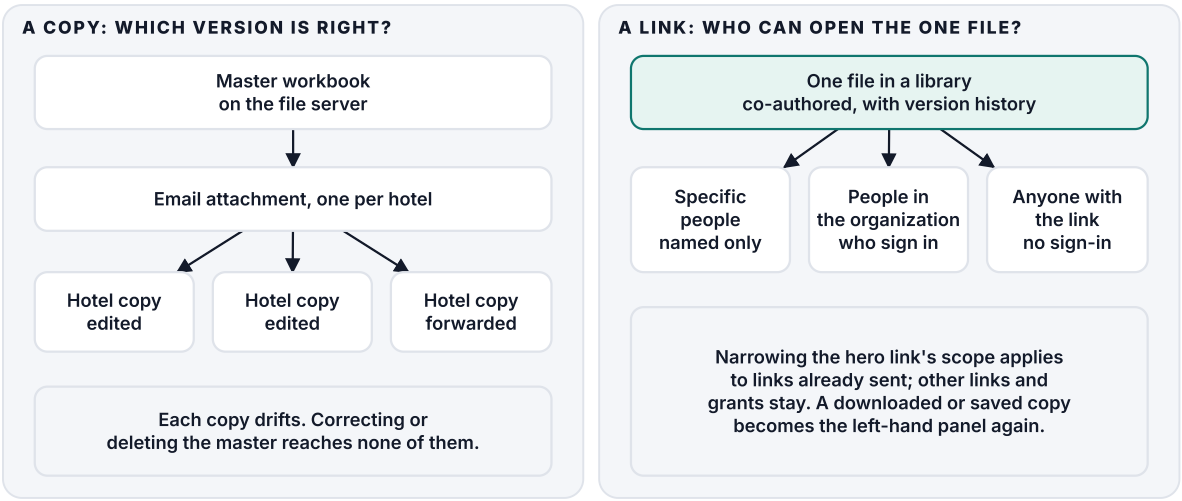
Link versus copy

For most of its life the group distributed the rate sheet as copies: faxed printouts, then floppy disks, then an email attachment to every hotel each season until its 2022 replacement. Each copy was a separate file. A correction at head office reached none of them, and the question every reconciliation asked was which copy was right.

Co-authoring changed the default. Microsoft's support pages say that "you and your colleagues can open and work on the same Excel workbook", and that co-authoring requires the workbook to be saved to OneDrive, OneDrive for Business or a SharePoint Online library; on-premises SharePoint sites do not support it. Several people edit one file, and the version history records the saves. Teams moved conversation onto the same files: a channel's Files tab is a folder in a SharePoint library. The working habit shifted from "send me the latest" to "send me the link".

FIGURE 9.1

Link versus copy: the question moves from which copy to who holds the link



A copy multiplies files that drift apart and that no change to the original reaches; a link keeps people on one file, so the control becomes who the link works for, and narrowing the hero link's scope applies to links already sent while other links and grants stay as they are.

Author's diagram based on Microsoft Learn, "How shareable links work in OneDrive and SharePoint in Microsoft 365" (ms.date 6 November 2026), and Microsoft Support, "Collaborate on Excel workbooks at the same time with co-authoring" (read 20 November 2026).

A link keeps people on one file, so while they stay on it the reconciliation problem largely goes away. It is replaced by an audience problem. The question becomes who holds the link, what scope it carries today, and which other paths reach the same file. That is a better question, because it has an answer the system can report. But the old question does not disappear. Every "Save a copy", every download, every file dragged from a library into an email starts a new copy with no link back, and Figure 9.1's right-hand panel becomes its left-hand panel again.

Migrations that widen access

A migration is a hop. It moves files, and it also rewrites who can open them, because the source and the destination describe access differently. The group's file server had its own history: NetWare trustee rights on the `S:\REVENUE` directory from 1994, remapped to Windows groups when the share moved to a Windows file server in 2006. By 2016 those groups included nested groups, departed staff and groups nobody could explain. The migration tool could resolve some of them to cloud identities and not others.

The team's choice is worth stating as it was made, not as a rule for everyone. Faced with a deadline and a list of unresolved groups, it mapped them all to one custom group, All staff (internal), and gave that group read access to the libraries that had held them. It did not record the decision as a permission change, did not ask the library owners, and did not compare the people in the old groups with the people in the new one. For most libraries the result was harmless. For the Revenue library it was not, because the workbook carried the CORP tab, with client contacts and negotiated rates, since 2004. Hotel general managers had received the tab by email every season; now every hotel and head office employee could open it. Files started a version history in the library, inherited its permissions, and none kept narrower ones. From then on the library held the file the seasonal email was sent from; the email to the hotels went on until 2022.

THE HOP · 5 · 2016 · WINDOWS FILE SERVER (GROUPS REMAPPED FROM NETWARE TRUSTEES IN 2006) TO SHAREPOINT ONLINE LIBRARY REVENUE

Record	The rate sheet (hotel and retail group), hop 5, 2016
From	Windows file server (groups remapped from NetWare trustees in 2006)
To	SharePoint Online library Revenue
Mechanism and schedule	A third-party migration tool moves the share to the cloud; to finish on time the team maps unresolved and nested groups to one custom group
Format change; what was lost or assumed	RIGHTS Files start a version history in the library; permissions inherited from the library; no item keeps narrower permissions. **The custom group All staff (internal), then the hotel and head office staff (about 4,000 people), gets read access to the Revenue library; the corporate tab becomes readable by all of them**
Accountable owner	Head of revenue (site owner)
Who can change it	The revenue team (edit); site collection administrators
Classification; retention or erasure duty	Personal data and commercial terms; no retention policy set on the library
Read by (people, apps, models, agents)	All staff (internal) (read): about 4,000 staff in 2016, few of whom need the CORP tab
Cost per month; basis	One-off: the migration project (not monthly); per month: a share of licences (illustrative)

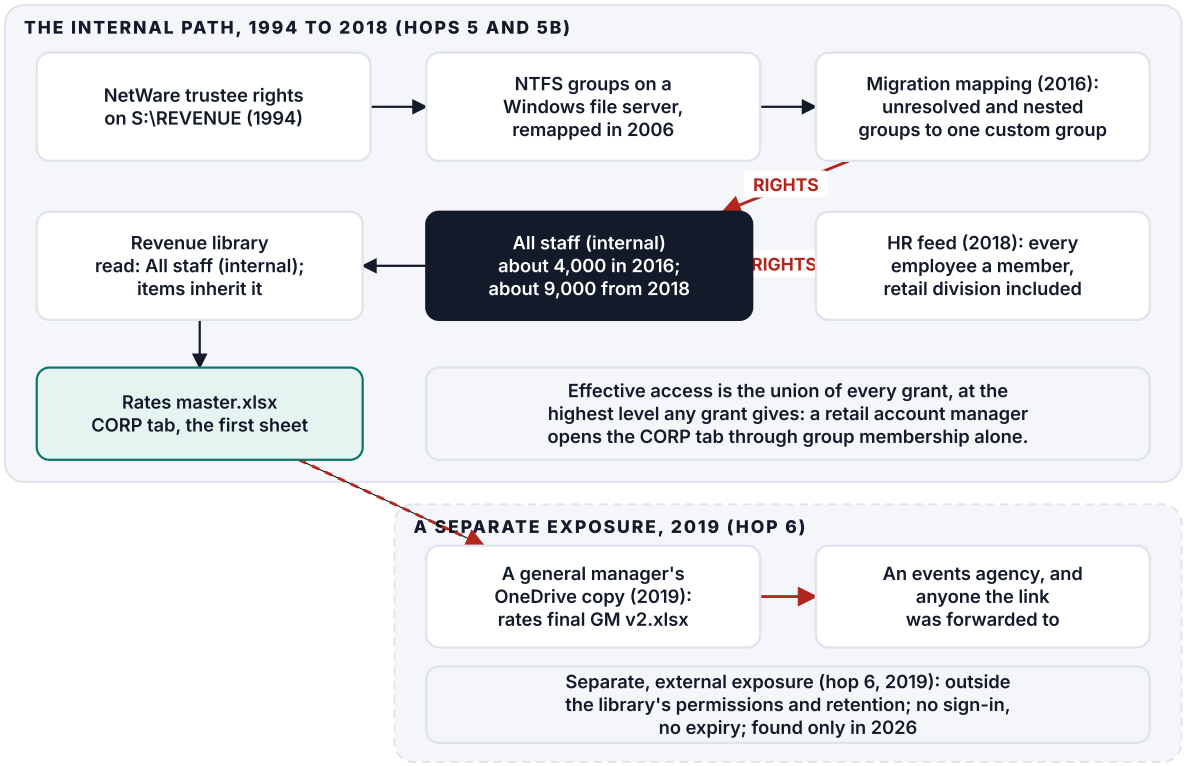
The custom group then did what groups do: it changed. In 2018 the retail division moved into the same tenant, and an HR feed made every employee a member of All staff (internal). The group grew from about 4,000 hotel and head office staff to about 9,000 people, including the retail division's account managers. Nobody reviewed what the group could reach. The Revenue library's read grant grew with it, without a single change to the library. This is **group drift**: a grant that was narrow enough when made becomes broad because the group's membership rule changed, and the owner of the resource is not told.

THE HOP · 5B · 2018 · HR SYSTEM TO THE GROUP ALL STAFF (INTERNAL)

Record	The group All staff (internal), the audience of the rate sheet, hop 5b, 2018
From	HR system
To	The group All staff (internal)
Mechanism and schedule	When the retail division moves into the same tenant, an HR feed makes every employee a member of All staff (internal); nobody reviews what the group can reach
Format change; what was lost or assumed	RIGHTS Dynamic group membership from the HR feed. **The group grows from hotel and head office staff to about 9,000 people, including the retail division's account managers; the Revenue library's read grant grows with it**
Accountable owner	Head of IT operations (identity)
Who can change it	The HR feed; identity administrators
Classification; retention or erasure duty	Group membership; no access review of what the group can reach
Read by (people, apps, models, agents)	Every member of the group, and every system that trusts it
Cost per month; basis	Per month: near zero; the cost is the reach (illustrative)

FIGURE 9.2

Effective access: how 9,000 people could open the corporate tab



No sharing link, and no single decision aimed at them, opened the workbook to the retail account managers: a 1994 trustee grant became Windows groups in 2006, a migration mapping gave one custom group read access in 2016, and an HR feed filled that group with every employee in 2018; the 2019 anyone-link copy is a separate, external exposure.

Author's diagram based on the book's running example (hops 5, 5b and 6; illustrative); link behaviour per Microsoft Learn, "How shareable links work in OneDrive and SharePoint in Microsoft 365" (ms.date 6 November 2026).

Figure 9.2 draws the path end to end. No link appears on the internal path. Every step was a reasonable decision by someone who could not see the whole: the 2006 remapping kept old grants, the 2016 mapping chose speed over review, and the 2018 feed served the identity team's purpose. The effective access that resulted belongs to nobody's design. Three controls would have caught it. Record the migration's permission mapping as a decision, signed by the owners of the libraries it touches. For each sensitive library, compare the people who could read it before and after the move, not only the group names. And review groups by what they can reach, so that a change to a membership rule triggers a look at every grant the group holds.

THE HOP · 6 · 2019 · REVENUE LIBRARY TO A GENERAL MANAGER'S ONEDRIVE COPY, SHARED WITH AN EVENTS AGENCY BY AN ANYONE-WITH-THE-LINK LINK

Record	The rate sheet (hotel and retail group), hop 6, 2019
From	Revenue library
To	A general manager's OneDrive copy, shared with an events agency by an anyone-with-the-link link
Mechanism and schedule	A general manager saves a copy to her OneDrive and sends the agency a link that needs no sign-in
Format change; what was lost or assumed	OWNERSHIP RIGHTS rates final GM v2.xlsx; a link with no expiry. **A copy outside the Revenue library's permissions; anyone holding the link can open it; the general manager, still employed, moves to another hotel in 2021 and the copy is never reviewed; found only in 2026**
Accountable owner	The general manager in name (personal drive); not reviewed after her 2021 move
Who can change it	The general manager; anyone with the link can view
Classification; retention or erasure duty	Personal data and commercial terms outside the library's controls
Read by (people, apps, models, agents)	The events agency and anyone the link was forwarded to (unknown)
Cost per month; basis	Per month: near zero (illustrative)

In 2019 a general manager at a city hotel saved a copy of the workbook from the Revenue library to her own OneDrive as `rates final GM v2.xlsx` and sent an events agency a link that needed no sign-in. The link had no expiry. The copy now sat outside the Revenue library's permissions: anyone holding the link could open it, and nobody at the group could say who that was. In 2021 she moved to another hotel. She was still employed, so no leaver process ever looked at her drive, and the copy was never reviewed. It was found only in 2026, during the walk back from the assistant's answer, and was then assessed as a possible personal data breach under the GDPR (Chapter Fifteen). The Ledger codes this hop OWNERSHIP and RIGHTS. The copy had an owner in name, the general manager, but nobody reviewed it after she changed hotels. And people the library never authorized could open it.

OUR VIEW, AS OF NOVEMBER 2026

In the case this book follows, the oversharing that mattered did not come from a link a user created. It came from a grant administrators made for a good reason and a group that grew afterwards. The pattern is worth checking for in any tenant. Before a workplace assistant retrieves from the libraries, the cheaper review is the group, not the file: list every group with read access to a library that holds personal or pricing data, count its members today, and ask the library's owner to confirm each one.

The owner when the author leaves

A cloud library or a personal drive needs an owner after its creator moves on, and the vendor's mechanisms differ for each. For a personal drive, Microsoft documents, as of its page dated 2 June 2026, that the OneDrive cleanup begins only when the user's account is deleted from Microsoft Entra ID; "no other action causes the cleanup process to occur, including blocking the user from signing in or removing the user's license". By default the user's manager is automatically given access to the OneDrive, a secondary owner can be named for users without a manager, and the default retention period for a deleted user's OneDrive is 30 days, which an administrator can set to any value from 30 through 3,650 days. If access delegation is off, or no manager or secondary owner is set, "no one has automatic access". Microsoft adds that retention policies and retention labels "always take precedence" over this standard deletion process, so content can be kept longer, or deleted sooner, than the OneDrive retention period. Separately, OneDrive accounts without a valid licence are archived on their 93rd unlicensed day.

That process is built for leavers. The general manager of hop 6 was a mover: she changed hotels and kept her account, her drive and her anyone link. No leaver step applies to a mover, so a mover review has to exist on its own: when a person changes role, the files they shared outward in the old role are listed and confirmed or removed.

Sites and groups have their own owners. Microsoft documents that a team or a Microsoft 365 group "can become ownerless if you delete an owner's account", and that an administrator can set a policy that emails the group's most active members and asks whether they accept ownership; if nobody accepts, the policy stops and administrators must assign an owner themselves. The policy applies only to Microsoft 365 groups. For SharePoint sites more widely, Microsoft describes site ownership policies, part of SharePoint Advanced Management, as of its page dated 14 May 2026: a policy sets a minimum number of owners, runs monthly in active mode, and can notify current owners, active members or the managers of previous owners. Microsoft notes that a departed user's details are deleted 30 days after they leave, so managers might receive only one notification, and that the stronger enforcement actions, read-only or archive, apply only to sites with no owner at all.

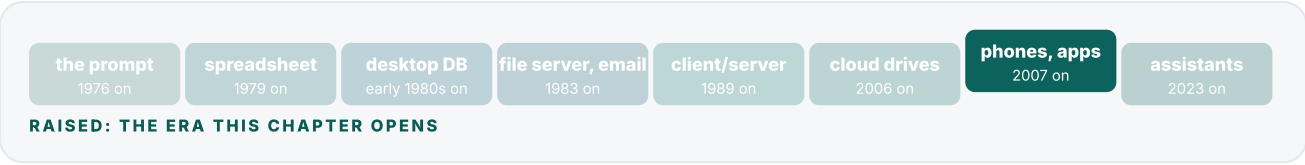
The group had neither. The head of revenue was the Revenue site's owner from the 2016 migration. In 2023 the site owner left and nobody was assigned. By then the workbook had been replaced by the revenue management system and retired by an emailed memo in July 2022, but the library still held it, All staff (internal) could still read it, and no one was responsible for either fact. Chapter Twelve tells the retirement; Chapter Fifteen tells the reassignment. The lesson for this chapter is plainer: **an owner is a field that must be filled again every time a person moves**, for libraries and sites as much as for files.

> **ERA CARD**
>
> **Reader.** Anyone a group or a link reaches: a colleague in another division, a partner holding a forwarded link, and, in the group's 2025 pilot, a workplace assistant that retrieves indexed content each user is permitted to access.
>
> **Security.** Inherited permissions, sharing links with three audiences, groups fed by HR systems; effective access is the highest grant from any source, and nobody sees the total unless they ask for it.
>
> **Cost.** Storage and version history are bundled with licences; the cost is the review nobody schedules, and the clean-up when a broad grant is found years later.
>
> **AI.** Microsoft documents that an item shared with specific people or security groups can appear in search and be "accessible through Copilot" for them, subject to their permissions (November 2026): a broad group is a broad audience for the assistant too.

Monday rule: for every library that holds personal or pricing data, list each group with access, its member count today and its membership rule, and ask the library's named owner to confirm or remove each grant; where no owner is named, assign one first.

CHAPTER TEN

The Copy That Left the Building



On a Friday evening a clinician at the health system reports her work phone missing after a train journey. The phone holds her email, with a spreadsheet of the clinic's waiting list attached to a message from the week before, and a messaging app the clinic uses to hand over patients. The privacy officer's first questions are not about the spreadsheet's contents. They are about the phone: was it encrypted, was it locked, was the passcode written anywhere, can it be wiped, and which other copies of the same list exist. Every one of those answers was decided months earlier, by whoever set the policy for devices and apps. This chapter follows the copy that leaves the building on a laptop, a USB drive, a sync cache or a phone, and the database that sits inside almost every app. The illustrative cases are the health system's lost phone and the insurer's loss adjusters, who work offline in the field; the rate sheet's journey adds no hop here. The decision it changes: **decide which records may leave on which devices, encrypted how and removable how, before the loss, because after it the only question is what the device already enforced**.

Four ways out, and two side doors

A file leaves the building in four common ways. A **laptop** carries whatever its user saved or synced. A **USB drive** or memory card carries a copy with no link to anything. A **sync cache** holds local copies of cloud files so they open without a connection. A **phone or tablet** holds attachments, downloads and the private storage of every app on it. Each is a copy in the sense of this book. The library's permissions, version history and retention settings do not follow a downloaded file or a USB copy. A managed sync cache is partly different, because the sync client and the user's sign-in still reach it, but anyone who can use the unlocked device can open what is cached. Figure 10.1 adds two side doors: the conflict copy a sync engine makes, and personal storage outside the organization. What governs a copy on a device is mostly the device's encryption, its lock, the app's policy and the user's habits.

The sync cache is the least visible of the four, because it looks like the cloud. Microsoft documents three states for OneDrive files on Windows. "Online-only files don't take up space on your computer" and cannot be opened without a connection. When a user opens one, "it downloads to your device and becomes a locally available file", which can be opened "even without Internet access". A file the user marks to always keep on the device stays there for offline use. A laptop that has opened the rate workbook once holds a local copy of it until the file is returned to online-only, whatever the library's settings say about who may read it.

The protections differ by platform, and this is a place where the vendors' documents really do differ.

On each platform. Windows: Microsoft documents BitLocker as encrypting "the OS, fixed data, and removable data drives (BitLocker To Go)", and a feature called device encryption that turns BitLocker on automatically on qualifying devices; device encryption "doesn't encrypt external/USB drives", and on a device that uses only local accounts it "remains unprotected even though the data is encrypted" (Microsoft Learn, November 2026). macOS: Apple documents that on a Mac with macOS 26.4 or later "FileVault is turned on by default", and that users can opt out (Apple Platform Security, published 3 August 2026). iOS and iPadOS: Apple documents that setting up a device passcode automatically enables Data Protection (Apple Platform Security, December 2024). Android: the Android Open Source Project documents that Android 7.0 and later support file-based encryption and that "full-disk encryption is not allowed on new devices running Android 10 and higher". Cloud: the library copy keeps its permissions, versions and retention; none of them follow the local copy.

Two gaps follow from that list. Encryption on a laptop does not by default extend to the USB drive plugged into it, and encryption that depends on a passcode protects little when the passcode is weak, and nothing once it is shared or taped to the device. Both are organizational rules, not product faults.

The database inside the app

When a phone app keeps a list of patients, claims or contacts for offline use, it often embeds a database library and keeps the data in a file, with no database server at all. A widely used example is SQLite. The project's own site describes it as "an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine", and says that "a complete SQL database with multiple tables, indices, triggers, and views, is contained in a single disk file". The code is in the public domain. The project's release history lists SQLite 1.0 on 17 August 2000.

The site also makes a claim about its reach, in these words: "SQLite is the most widely deployed database in the world with more applications than we can count". A separate page lists where it runs, including every Android device, every iPhone and iOS device, every Mac, every Windows 10 and 11 installation, and the Firefox, Chrome and Safari browsers, and it adds its own caution: "Precise numbers are difficult to obtain and so exact rankings are impossible." This book takes the list as the project's statement and draws one conclusion from it: if the project's list is even roughly right, the devices that carry your records already hold database files, inside apps you did not write.

Three properties of that file matter for a copy census.

- **It is portable.** SQLite's site says the file format is cross-platform: "you can freely copy a database between 32-bit and 64-bit systems or between big-endian and little-endian architectures". A database copied off a device, with the companion files below, is a readable copy of the app's data, tables and all.
- **It has companions.** SQLite documents temporary files that sit "in the same directory as the database file": a rollback journal named with "-journal" appended, usually deleted when a transaction ends, and in write-ahead log mode a file named with "-wal" appended, which "will remain in the filesystem" if the last connection does not shut down cleanly, plus a "-shm" shared-memory file. A copy of the app's storage may include recent changes in these files that are not yet in the main one.
- **A delete may leave traces.** SQLite documents that its secure_delete setting "is normally off", and that "applications that wish to avoid leaving forensic traces after content is deleted or updated should enable the secure_delete pragma prior to performing the delete or update, or else run VACUUM after the delete or update". A record deleted inside an app is gone from the app's screens; whether its bytes are gone from the file depends on how the app was built.

None of this makes SQLite a risk. It is a well-documented engine doing what its documentation says. The point is for the inventory: a phone app that works offline holds a database, and that database is a copy of whatever the app synced. It belongs in the census with its owner, the organization that deployed the app, and its control, the policy on the app.

Offline copies and sync conflicts

The insurer's loss adjusters show the other side of offline work (an illustrative case). After a storm, adjusters visit damaged homes in areas without a mobile signal. Their tablet app stores claim notes, measurements and photos locally and uploads them when the tablet reconnects. That design is right: an adjuster who cannot record a visit until the network returns does not record it well. But it means that for hours or days the record of the visit exists only on the tablet, and that two people can change the same claim in two places.

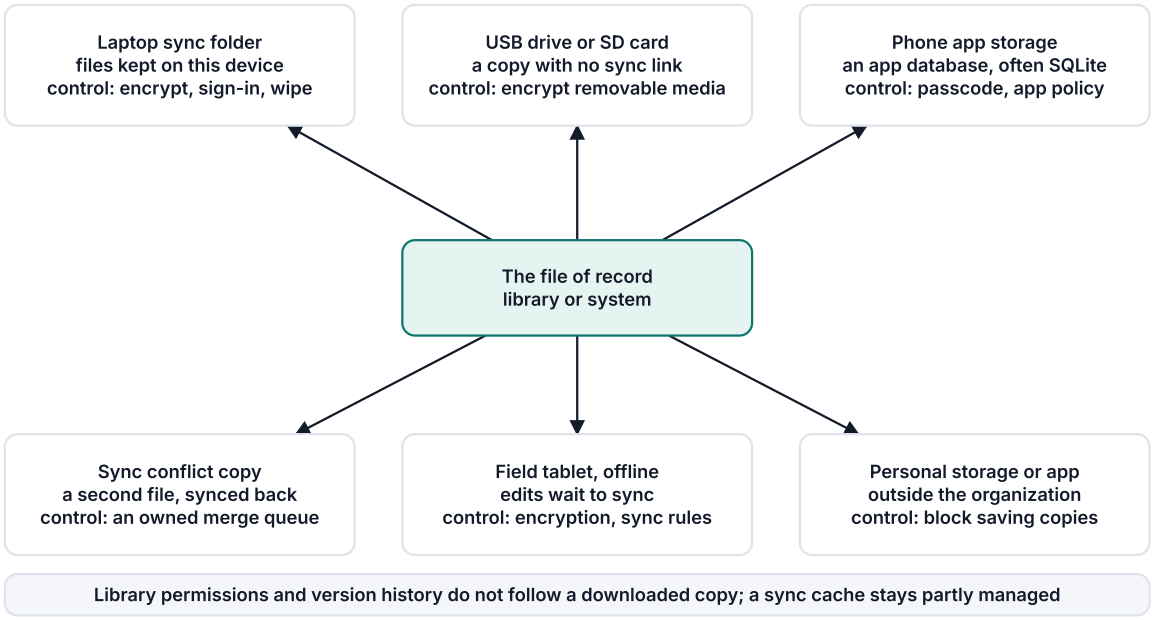
When the copies meet again, a sync engine has to decide. The adjusters' app has its own rules; OneDrive for work or school, which Microsoft documents, shows the usual choices. For most Office files, the conflict opens an experience in the Office app with two choices: "Save a Copy" outside the synced folder, or "Discard" the local changes and fetch the server version. For other file types, "OneDrive automatically keeps both versions": the online version keeps the original name, and "the copy on your computer has your device name appended to the file name such as Report-JOHNS-SURFACE.txt". Microsoft notes that OneDrive for work or school "creates up to 5 conflict versions for non-Office file types", and leaves the decision to the user: compare and merge, rename, or keep both. For co-authored workbooks, Microsoft's support pages say that changes made offline are not merged until the device reconnects, and that when people change the same content, generally the last change saved "wins".

Each of those outcomes is reasonable, and each one can create a copy or a silent overwrite. A conflict copy with a device name in it is a new file in the library with no owner and no stated meaning. A last-write-wins merge replaces one adjuster's measurement with another's and records only that a save happened. The claims tracker of Chapter Five, still running in 2026, receives what the tablets upload; it does not know which tablet's version of a claim was the one the adjuster meant.

The controls are operational. Name which system holds the record when two copies disagree, and make the app say so on screen. Give conflict copies an owner: a queue that someone empties every day, not a folder nobody opens. And keep corporate data inside managed apps. Microsoft documents that Intune app protection policies "ensure an organization's data remains safe or contained in a managed app", that they can apply on devices "not enrolled in any mobile device management solution", typically employee-owned, and that a selective wipe "removes company app data from an app" without wiping the device. Settings such as "Save copies of org data" control where a user can save a work file from a protected app (Microsoft Learn, page dated 4 March 2026).

FIGURE 10.1

Where a copy goes when it leaves the building, and what still governs it



The four ways out (a laptop and its sync cache, removable media, and a phone or tablet, drawn here as a phone app and a field tablet) and two side doors (a sync conflict copy and personal storage): the library's permissions and version history do not follow a downloaded copy, and what governs each one is mostly the device, the app and the policy on them.

Author's diagram based on Microsoft Support, OneDrive Files On-Demand; Microsoft Learn, "Resolve sync issues in OneDrive for work or school", "BitLocker device encryption" and "App protection policies overview"; sqlite.org, "About SQLite" (all read 20 November 2026).

The lost phone and the breach rule

Back to the health system's clinician. The waiting list on her phone holds names, dates of birth and the clinic each patient is waiting for: protected health information under the HIPAA rules. What the health system owes after the loss is set by the Breach Notification Rule, 45 CFR 164.400 to 164.414, which applies to breaches occurring on or after 23 September 2009.

The rule turns on two definitions in 45 CFR 164.402. A breach is an acquisition, access, use or disclosure not permitted by the Privacy Rule "which compromises the security or privacy of the protected health information", and outside three listed exceptions it "is presumed to be a breach unless" the covered entity "demonstrates that there is a low probability that the protected health information has been compromised based on a risk assessment" of at least four factors: the nature and extent of the information, including the types of identifiers and

the likelihood of re-identification; the unauthorized person who used it or to whom it was disclosed; whether it was actually acquired or viewed; and the extent to which the risk has been mitigated. The notification duties apply to **unsecured** protected health information, defined as information "not rendered unusable, unreadable, or indecipherable to unauthorized persons through the use of a technology or methodology specified by the Secretary" in guidance.

That guidance is what practitioners call the encryption safe harbor. The Department of Health and Human Services specifies encryption and destruction as the methods. Electronic information counts as encrypted when it meets the Security Rule's definition of encryption and "such confidential process or key that might enable decryption has not been breached"; HHS advises that decryption tools "should be stored on a device or at a location separate from the data", and that valid encryption for data at rest is consistent with NIST Special Publication 800-111. HHS states that entities that secure information as the guidance specifies "are relieved from providing notifications following the breach of such information".

So the answer for the lost phone is an assessment, not a rule of thumb. If the phone's storage was encrypted as the guidance describes and the means to decrypt it, here the passcode that unlocks the device's keys, was not compromised, the information may count as secured under the HHS guidance, and in that case the notification duties do not apply. If the phone was unlocked when lost, or the passcode was written on its case, the safe harbor is in doubt, the presumption of breach applies and the four-factor assessment decides. An unencrypted copy of the same list elsewhere does not change the answer for the phone, but if that copy was also exposed, it is a second event to assess. Neither outcome follows from the word "encrypted" alone, and the burden falls on the health system: 45 CFR 164.414 places on the covered entity "the burden of demonstrating that all notifications were made as required by this subpart or that the use or disclosure did not constitute a breach".

If notification is due, the clock is fixed in the rule. A breach is treated as discovered on the first day it is known, or would have been known with reasonable diligence, to any workforce member or agent of the covered entity other than the person who committed it. Individuals must be notified "without unreasonable delay and in no case later than 60 calendar days after discovery". A breach involving more than 500 residents of a State or jurisdiction also requires notice to prominent media outlets there. The Secretary is notified contemporaneously with individuals for breaches of 500 or more people; smaller breaches go in a log, reported within 60 days after the end of the calendar year.

The health system's privacy officer can make that assessment quickly only if three facts are already on record: the device's encryption state and management status, whether it could be wiped and was, and every other copy of the list. The first two belong to the device policy. The

third is this book's subject. A waiting list emailed to a phone is also in a mailbox, perhaps on a laptop, and maybe in the clinic's messaging app; the phone is one row of its copy census.

ERA CARD

Reader. Whoever holds the device: the employee, a family member, a finder, a thief; and every app on it that can read shared storage.

Security. Device encryption, passcodes, app protection policies and remote or selective wipe; the library's permissions stay behind when the copy leaves.

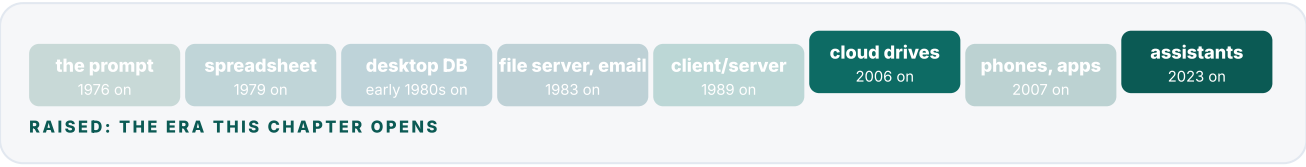
Cost. Devices and management licences are budgeted; the costly part is the assessment after a loss, and the people hours to find every other copy.

AI. An assistant drafting on a phone works on the same local files and attachments; a summary saved or pasted on the device is one more copy for the census.

Monday rule: pick your three most sensitive desk files, list every device class that can hold a copy of each, and confirm for each class that storage encryption, a wipe path and an app policy are in force, writing down where they are not.

CHAPTER ELEVEN

Deleting a File for Real



In 2020 a guest of the harbor hotel, a regular who appeared in the rate sheet's VIP tab, asked the hotel and retail group to erase her personal data. The General Data Protection Regulation had applied since 25 May 2018, and its Article 17 gave her that right on the grounds it lists. The privacy team removed her from the CRM, opened the current workbook in the Revenue library, deleted her row and closed the request in about two hours. In 2026 the incident team found her name again, in a OneDrive copy that an anyone link still opened without sign-in. Nobody had lied, and the team had followed the practice of the day. The deletion simply reached the two places the team knew about. This chapter is about what delete means once a record lives in a cloud library, a mailbox, a backup and a search index at the same time. The decision it changes: **treat every deletion as a hop with a reach, and record which copies it reached and which it did not**.

What delete means in a library

In a cloud library, delete is a move. Microsoft documents that a file deleted from a SharePoint or OneDrive site goes to the site's first-stage recycle bin, which the user can see and restore from; if the user empties it, the file moves to a second-stage recycle bin that only site collection administrators can see. Microsoft states that "a 93-day retention period spans both the first- and second-stage recycle bins", after which the file is permanently deleted from wherever it sits (as of 20 November 2026). For 93 days, then, a deleted workbook is still a workbook.

A file in a library is also more than its current version. Microsoft documents that, by default, versioning in SharePoint and OneDrive libraries "retains a minimum of 500 major versions", a limit an administrator can change. The Revenue library has kept versions of Rates master.xlsx since the 2016 migration. Deleting a row from the current file, as the privacy team did, changes the next version. It does not rewrite the earlier ones: every version

saved while the row existed still holds it, unless a version limit has already trimmed that version. Version history is not a backup either: it lives in the same library, under the same permissions.

Mail follows a similar pattern. Microsoft documents that an item a user deletes from Deleted Items in Exchange Online moves to a hidden Recoverable Items folder, where "the default deleted item retention period for Exchange Online is 14 days", which an administrator can raise "up to a maximum of 30 days". The seasonal emails that carried the rate sheet to the hotels since 1998, and the CORP and VIP tabs since 2004, sit in dozens of mailboxes. Each mailbox follows its own clock.

Retention and legal hold: when delete is not allowed

An organization can also tell the platform to keep content whatever a user does. In Microsoft Purview, a retention policy or retention label can keep files for a set period. Microsoft documents that when a user deletes an item subject to retention, or edits one under a retention policy, "the original content is copied to the Preservation Hold library", a hidden location in the site that users are not meant to open; compliance staff reach it with eDiscovery tools. A timer job that runs every seven days removes content that has sat there for more than 30 days and whose retention period has expired; Microsoft notes that, with that 30-day minimum, "it can take up to 37 days for content to be deleted from the Preservation Hold library", and that the content then passes through the second-stage recycle bin.

Two of Microsoft's rules decide what happens when settings conflict: "retention always takes precedence over permanent deletion, and the longest retention period wins". A legal hold works the same way. Microsoft documents that permanent deletion "is always suspended" if the item must be retained under another policy or label, "or it's under eDiscovery holds for legal or investigative reasons", and that for items under a retention policy or an eDiscovery hold the library's version limits are ignored, so old versions are not purged. This is the right behavior for a regulator or a court. It also means that a privacy team can delete a file and the platform can keep it, correctly, for years. The person who deletes and the person who set the hold rarely talk.

Search follows the same split. Microsoft documents that "the Recycle Bin isn't indexed and therefore unavailable for searching" (as of 20 November 2026). A deleted file leaves organization-wide search, and with it the reach of an assistant that retrieves through that search, while it still sits in the recycle bin for an administrator to restore. Copies kept for retention remain searchable by eDiscovery for compliance staff. Deleting the file leaves three groups with different views: users cannot find it in search, though they can restore it from

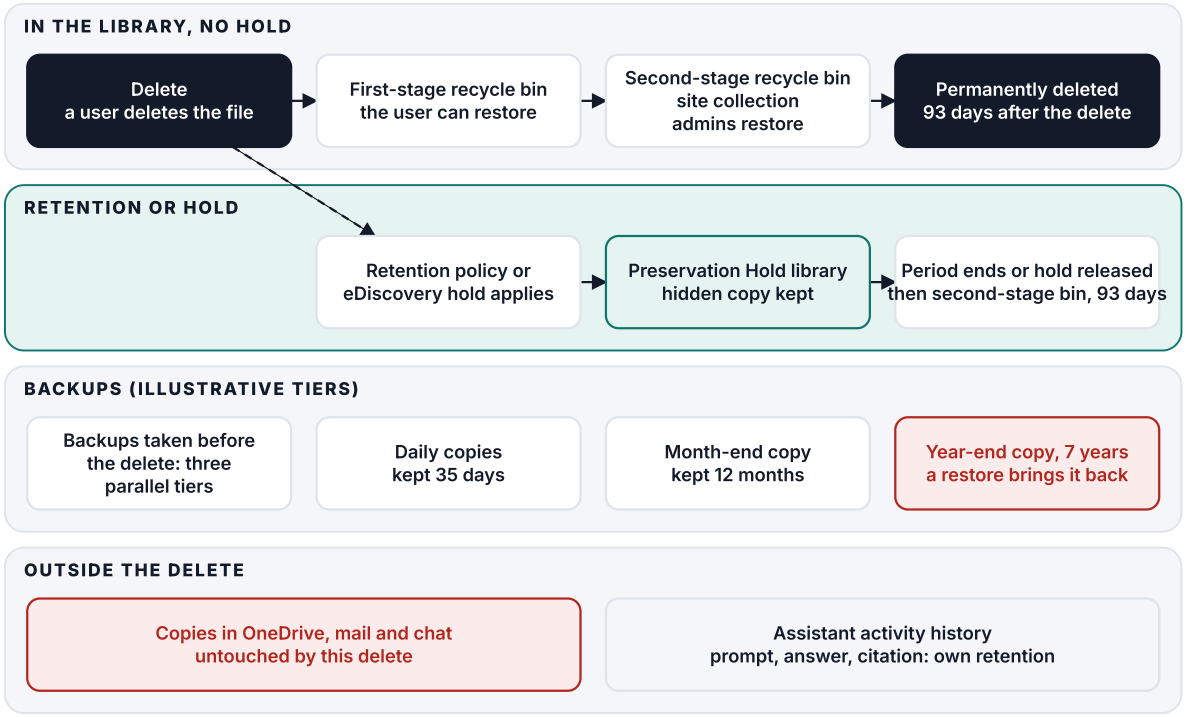
the first-stage bin; site collection administrators can restore it from either bin within the 93 days; and compliance staff can search any retained copy. Copies elsewhere, in mailboxes, OneDrive drives and backups, are unchanged, because the delete never touched them.

The assistant adds a store of its own. Microsoft documents that when a user works with Microsoft 365 Copilot, the stored data "includes the user's prompt and Copilot's response, including citations to any information used to ground Copilot's response", and that this data sits in "a hidden folder in the mailbox of the user". Microsoft lists two events that remove those messages: a user deleting the chat, and a request that deletes the user's activity history; a retention policy that expires removes them by a separate path, through a hold area and then deletion (as of 20 November 2026). Deletion of a cited file is not on that list. An answer that quoted a row can outlive the row.

Figure 11.1 draws the life of one deleted file in the group's tenant, with the legal hold branch and the backup branch.

FIGURE 11.1

The life of a deleted file: 93 days in the recycle bins, longer under a hold, years in a backup



A file deleted in a SharePoint library passes through two recycle bins for 93 days in total, is kept in the Preservation Hold library while a retention policy or legal hold applies, and survives in backups until the longest backup tier that captured it expires.

Author's diagram based on Microsoft Learn, "Learn about retention for SharePoint and OneDrive" and "Learn about retention for Copilot and AI apps" (read 20 November 2026); backup tiers are illustrative.

Backups and the retention arithmetic

Backups sit outside the library's recycle bins, versions and holds. Whatever a team runs, a backup is a copy taken on a schedule and kept by tiers, and the arithmetic of those tiers decides how long a deleted row survives. A worked example, illustrative and not the group's real policy: daily backups are kept for 35 days, a month-end backup for 12 months and a year-end backup for seven years, each tier a point-in-time copy.

Suppose the row is deleted on 10 March. The daily copies taken before that day expire one by one, the last on about 14 April. The month-end copy of 28 February holds the row until the end of February the following year. The year-end copy of 31 December holds it for seven years. So the row's real end date is set by the longest tier that captured it, not by the day it

was deleted. If the row was added and deleted between two month-end copies, only daily copies hold it, and it is gone after about five weeks. If it sat in the workbook for 16 years, as her row did from 2004 to 2020, every tier holds it.

Backup tier (illustrative)	Kept for	Row deleted on 10 March holds until
Daily copies before 10 March	35 days	about 14 April
Month-end copy of 28 February	12 months	end of February next year
Year-end copy of 31 December	7 years	seven years after that 31 December

A restore makes the arithmetic matter. A library restored from the month-end copy brings back every row that copy held, including a row erased since, unless someone reapplies the erasures. A list of completed erasure requests, kept apart from the data and checked after each restore, is the control. No single rule says how far an erasure reaches into backups; each regulator gives its own guidance. The UK Information Commissioner's Office, writing for UK law, says "the key issue is to put the backup data 'beyond use', even if it cannot be immediately overwritten", and that organizations "must be absolutely clear with individuals as to what will happen to their data", backups included. The harbor hotel answers to an EU supervisory authority, so that guidance is an example of regulatory thinking, not the rule the group follows.

Erasure and media sanitization

Article 17 of the GDPR gives a person "the right to obtain from the controller the erasure of personal data concerning him or her without undue delay" where one of its grounds applies, for example that the data is no longer needed for the purpose it was collected for, or that consent is withdrawn. Article 17(3) lists exceptions, among them processing needed for freedom of expression and information, for compliance with a legal obligation, for public health, for archiving in the public interest, research or statistics, or "for the establishment, exercise or defence of legal claims". A legal hold can support that last exception, but only where the controller records why the data is needed for a claim; the hold alone does not establish it. Article 12(3) sets the clock: the controller answers "without undue delay and in any event within one month of receipt of the request", extendable "by two further months where necessary". None of this says every backup must be rewritten at once; it says the controller must know where the data is and decide, for each place, what erasure means and record that decision.

Media sanitization is a separate duty, about the device, not the record. NIST Special Publication 800-88 Revision 2, *Guidelines for Media Sanitization*, published in September 2025, replaced Revision 1 of December 2014. It describes three methods. Clear protects "against simple, non-invasive data recovery techniques using the same interface that is available to the user", for example by overwriting. Purge makes recovery "infeasible using state-of-the-art laboratory techniques" and leaves the device "in a potentially reusable state". Destroy does the same and leaves the device unusable for storage. Rev. 2 also warns that on cloud storage, where the data owner has no direct access to the physical disk, "cryptographic erase may be the only viable purge sanitization technique option". For desk files, the two duties meet when a laptop or a USB stick carrying a copy is retired: the copy must be counted before the device is sanitized, or the count is lost with it.

Archive is not deletion, and deletion is a hop

Archiving moves a file somewhere cheaper or quieter. It does not end the duty attached to the data. Microsoft documents that sites moved to Microsoft 365 Archive remain subject to the retention policies and labels that covered them; an archived copy of the rate sheet still holds the VIP rows and still falls under an erasure request. The difference matters in Chapter Fifteen, where the group takes the workbook out of search by removing access, not by archiving it (Microsoft documents that archived content keeps end-user search): removing access takes it out of the assistant's reach, which is a retrieval control, while the erasure duty is met separately, copy by copy.

The Ledger treats a deletion like any other hop. A deletion row reads: from the request or retention rule, to the copies it reached, with the copies it missed in the loss column. That is why hop 6b carries the RIGHTS code. The erasure duty is a legal requirement attached to the data, and it stayed behind when the data had already moved.

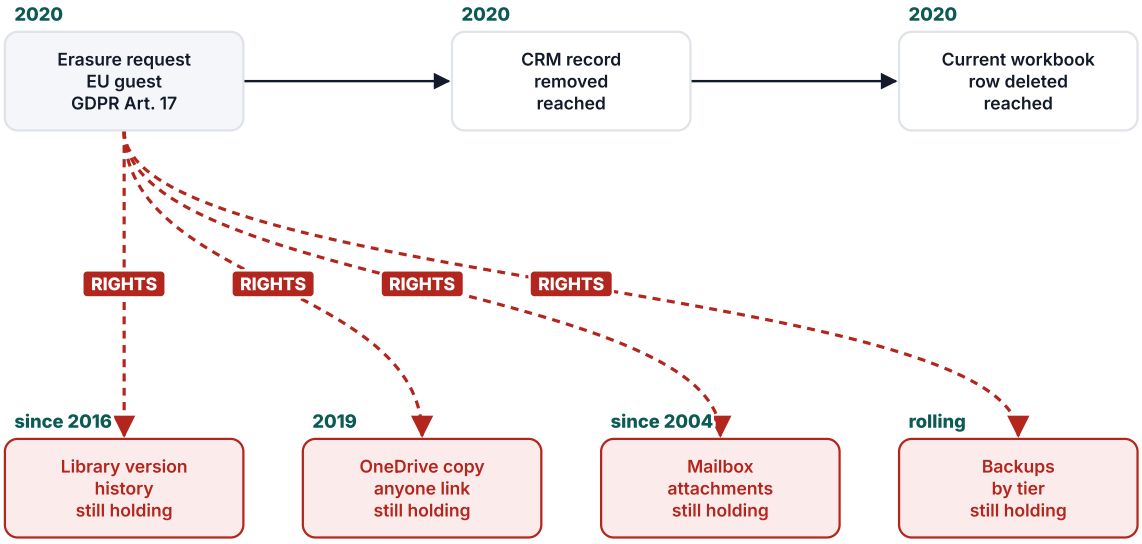
THE HOP · 6B · 2020 · ERASURE REQUEST FROM AN EU GUEST (GDPR ART. 17) TO THE CRM AND THE LIBRARY'S CURRENT FILE

Record	An EU guest's row in the rate sheet (erasure request), hop 6b, 2020
From	Erasure request from an EU guest (GDPR Art. 17)
To	The CRM and the library's current file
Mechanism and schedule	The privacy team removes the guest from the CRM and deletes the row from the current workbook
Format change; what was lost or assumed	RIGHTS Deletion in two places. **The erasure duty does not travel: the guest stays in the library's version history, the 2019 OneDrive copy, mailbox attachments since 2004 and backups**
Accountable owner	Privacy team
Who can change it	The privacy team; the revenue team
Classification; retention or erasure duty	Personal data under an erasure request
Read by (people, apps, models, agents)	Anyone who can open an old version or a copy
Cost per month; basis	One-off: 2 hours of work (not monthly); the exposure stays open (illustrative)

Hop 6b is a side branch of the thread. In 2020 the privacy team received an erasure request from an EU guest under GDPR Article 17. It removed her from the CRM and deleted her row from the current workbook in the Revenue library, about two hours of work. The request reached one copy of the workbook. The guest stayed in the library's version history, in the 2019 copy in a general manager's OneDrive that an anyone link still opened, in mailbox attachments going back to 2004 and in backups. Anyone who could open an old version or a copy could still read her row. Figure 11.2 shows the reach. The next chapter goes looking for every copy.

FIGURE 11.2

The erasure that cleared two places and missed four that still held the guest



Red arrows and chips mark a hop where something was lost (loss codes, Chapter 2).

In 2020 the privacy team cleared the CRM and the library's current file; the guest's row stayed in version history, the 2019 OneDrive copy, mailbox attachments since 2004 and backups, which is the RIGHTS loss of hop 6b.

Author's diagram based on the book's running example; the hotel and retail group, its people and numbers are illustrative.

ERA CARD

Reader. The privacy team and the records manager, who must now know where a record lives before they can say it is gone; the assistant, which reads whatever a deletion missed.

Security. Recycle bins, version history, retention policies and legal holds keep data on purpose; the risk is a deletion that nobody checked against them.

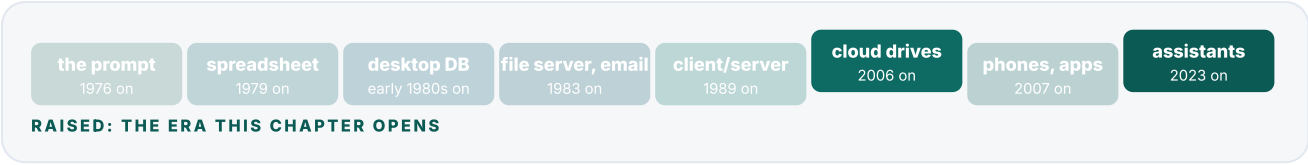
Cost. Storage for retained versions and backups is small; the cost is the search for copies after a request, and a restore that brings back what was erased.

AI. Microsoft documents that assistant interactions, with their citations, are stored and deleted on their own schedule, so an answer can outlive the file it quoted.

Monday rule: for the next erasure request, record every place the data was found, what was done in each (deleted, held under a named exception, put beyond use) and who checked it, and keep that list where a backup restore will consult it.

CHAPTER TWELVE

Shadow Data: Finding Every Copy



On Thursday 12 March 2026, once the revenue manager had traced the assistant's answer to Rates master.xlsx, the incident team asked a question nobody in the group had asked in 39 years: how many copies of the rate sheet exist, and who can open each one? They searched mailboxes, exports, version histories, OneDrive copies and the assistant's index for the workbook and the personal data in it. The census found 23 copies, counted by place and with library versions excluded; three sat outside the group. None of the 23 carried a readable record of whether it was current, retired or a working draft; the only trace of a state was an emailed memo about one of them. The two right sources, the revenue management system and the CRM, were counted separately. This chapter is about the copies nobody lists, how to find them and where each method stops. The decision it changes: **count the copies of an important record before you clean, retire or switch an assistant on, and record the limits of the count**.

Where copies live that nobody lists

Shadow data, in this book, means copies of a record that exist outside any list someone keeps. Many were made for a good reason and then forgotten. The places they collect are few, and each has a documented behavior.

- **Imports and exports of rented systems.** A system import is a copy taken on a date. The revenue management system kept the March 2022 import file in its history; the booking engine received one CSV per hotel each season from 2003. An export made to check an import is a copy too.
- **Mailbox attachments.** The rate sheet was emailed to the hotels each season from 1998 until 2022, and some hotels kept forwarding leftover copies after that. Each mailbox keeps its attachments until someone or some policy deletes them.

- **Chat files and sync caches.** Microsoft documents that files uploaded "in a one-on-one or group chat are stored in the Microsoft Teams Chat Files" folder of the uploader's OneDrive. Its sync client can mark files "Always keep on this device", and those files "download to your device and take up space" (as of 20 November 2026). A file uploaded to a chat is a new copy in a personal drive; a synced file is a copy on a laptop.
- **Hidden sheets.** A worksheet can be out of sight and still in the file. Microsoft's Excel documentation distinguishes a sheet hidden so that "the user can unhide via menu" from one set to "very hidden", which a user cannot unhide from the menu. Neither changes what a recipient of the file receives.
- **Chat and prompt histories.** Microsoft documents that Microsoft 365 Copilot stores the user's prompt and its response with citations, in a hidden mailbox folder that compliance staff can search with eDiscovery. A consumer chat service used on a personal account keeps its history under that service's terms, outside the organization's retention and contracts.
- **Backups.** Copies in the tenant can also sit in whatever backups the organization runs, each tier on its own clock, as Chapter Eleven showed.

The sheet nobody saw

SOURCED FAILURE

Police Service of Northern Ireland, 8 August 2023. The UK Information Commissioner's penalty notice of 26 September 2024 sets out the facts (paragraphs 6, 7 and 21). On 3 August 2023 the PSNI received a freedom of information request, through the WhatDoTheyKnow website, for the number of officers and staff at each rank and grade. A member of the workforce planning team prepared the figures in an Excel workbook built from a download of the human resources system: a worksheet titled "SAP DOWNLOAD" holding, among other fields, surnames and first-name initials, rank or grade, department, location and staff number for every officer and member of staff. Other worksheets were added for the analysis, one of them the "Return worksheet" with the answer. The team member then deleted all the tabs visible on screen except the Return worksheet. The source worksheet stayed in the file, out of view, signalled only by three dots beside the remaining tab (footnote 18). The quality check looked at the visible tab. At 14:31 on 8 August the response and the workbook were uploaded to the website, and the source worksheet was publicly accessible for about two hours and 20 minutes, until 16:51. It disclosed the personal data of 9,483 officers and staff.

The notice records that before the incident the PSNI's training did not raise "the risk that FOI response letters and their attachments might contain hidden data" and that there was no guidance on such checks (paragraph 44). It notes that it is "particularly easy for spreadsheet files (such as Excel files) to contain hidden data", including deliberately hidden sheets and the data embedded behind a pivot table (page 35). The Commissioner set a penalty of 750,000 pounds, reduced from 5.6 million under the approach for public sector bodies. The independent review that the PSNI and the Northern Ireland Policing Board commissioned, *Protecting From Within* (published 11 December 2023), describes the worksheet as "a hidden tab", says the incident "is considered to have been the most significant data breach that has ever occurred in the history of UK policing" and makes 37 recommendations on information management and security.

The rate sheet carries a milder form of the same risk. Its CORP tab is the workbook's first sheet, a small table of corporate rates and client contacts, and its VIP tab, pasted in 2004, sits further along the tab bar: visible, but where a general manager opening the file for the season's rates had no reason to look. Every seasonal attachment carried both. A check of what a reader looks at sees the rates and the CORP tab; the file also holds the guests. The PSNI's own interim fix, the notice records (paragraph 47), was to share "flattened PDF/CSV files only" in public responses: a copy that carries only what the sender meant to send.

Discovery methods and their limits

No single method finds every copy. Each one sees a slice, and the honest output of a search is a count plus a statement of what the count could not see. The table lists the methods the group used in March 2026; it names a category of tools, not products, and ranks none of them.

Method	What it finds	Where it stops
Search and content inventory	Files by name or content in indexed sites, drives and mailboxes	Only what is indexed and what the searcher may open; a renamed copy with edited content can slip through. Microsoft documents that the recycle bin "isn't indexed"
Audit logs	Who opened, shared or downloaded a file, and when	Retention: Microsoft documents 180 days by default for Audit (Standard) records created from 17 October 2023, one year by default for some record types of users licensed for Audit (Premium), up to ten years with an add-on. Older access is not recorded
eDiscovery	Mailboxes, OneDrive and SharePoint content, including hidden folders that hold assistant interactions and retained copies	Only the tenant; not personal accounts, devices or rented systems; Microsoft documents that a search finds no recycle bin content on which to place a hold
SaaS inventory	Which rented services received imports or exports	Lists services, not the files inside them; depends on what each contract lets the group export
Data loss prevention reports	Files and messages that match defined patterns or labels	Finds only what a rule describes; a rate table with no matching pattern passes
Endpoint and sync telemetry	Local copies, sync clients, removable media	Managed devices only
Data security posture tools (a category)	Many stores scanned, classified and mapped to who can access them	Coverage set by connectors and permissions granted to the tool; no proof of completeness
Interviews	Copies people know they made: the 2019 link, the 2024 chat	Memory, staff who have left, reluctance to admit a personal account

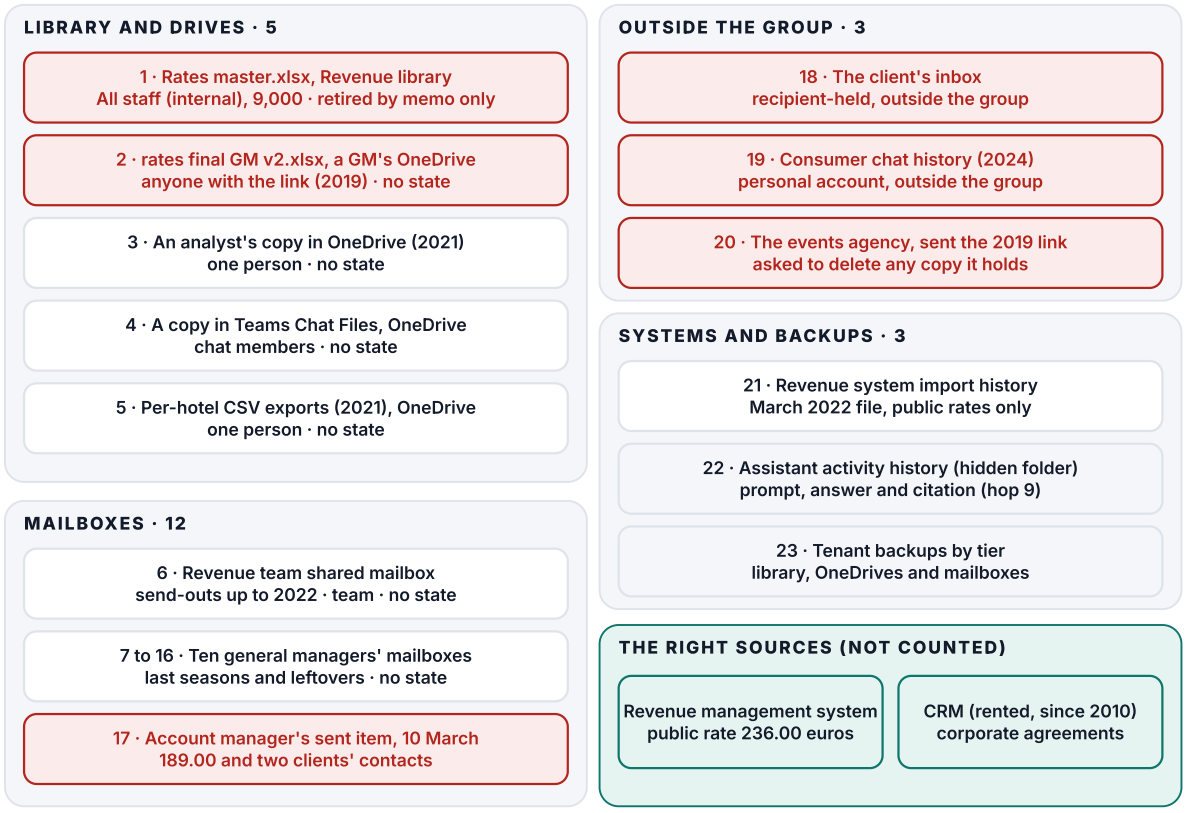
The audit row decided one of the group's answers. The Revenue library had been readable by All staff (internal) since 2016, but the audit records reached back only months, so the group could not prove who had opened the workbook in those years. The incident record says so in plain words, which is the correct output of a method that stops.

The 2026 copy census

The census needs a unit, and it counts places, not files. A copy is one place where the workbook, or rows taken from it, can be opened today: a library item, a file or one folder of exports in a drive, a mailbox however many attachments it holds, a history or a backup set. Library versions are not counted, because they belong to the item; nor is the search index, which is a view of the library item, not a separate copy. The assistant's activity history counts apart from the mail in the same mailbox, because it has its own retention and is searched with different tools. The revenue system's import history counts because the group can open it today; rented systems that received uploads in earlier years, such as the booking engine, are recorded as limits of the count until their vendors confirm what they still hold. Each copy gets three facts: its location, its sharing scope (one person, a group, everyone in a group of about 9,000, anyone with a link, or outside the group) and its authority state. Figure 12.1 shows the result: 23 copies (illustrative). Five sit in the library and personal drives, twelve in mailboxes, three outside the group, and three in systems and backups. Only one has anything like an authority state, and that state lives in an email: the library file was retired by a memo in July 2022. The other 22 have none. The two right sources sit beside the census and outside the count: the revenue management system for public rates and the CRM for corporate agreements.

FIGURE 12.1

The 2026 copy census: 23 copies of one workbook, one retired only by an emailed memo, none marked in the file, and the right sources outside the count



Each copy is one place where the workbook, or rows from it, can be opened in March 2026, with its sharing scope and authority state; red marks broad or external reach and the incident's own copies, and the two right sources sit outside the count.

Author's diagram based on the book's running example; the census, the hotel and retail group and its numbers are illustrative. Library versions and the search index are not counted.

Two copies explain most of the 2026 incident. Copy 1, the library item, is the one the assistant read. Copy 2, the general manager's OneDrive copy behind an anyone link, is the separate external exposure, and it still held the VIP tab with the guest erased in 2020. The rest explain the cost of the clean-up: every mailbox, export and history is a place where the incident team had to decide whether to delete, hold under a stated exception or restrict.

A retirement only an email knew, and a paste nobody saw

Two hops of the thread created copies the census had to find: the retirement that did not retire anything, and a copy that left the group through a personal account.

THE HOP · 7 · 2022 · RATES MASTER.XLSX TO RENTED REVENUE MANAGEMENT SYSTEM (ONE IMPORT)

Record	The rate sheet (hotel and retail group), hop 7, 2022
From	Rates master.xlsx
To	Rented revenue management system (one import)
Mechanism and schedule	Public rates are imported in March 2022 and then maintained in the new system (236.00 euros: the harbor hotel's public rate for the 2026 festival dates); the workbook is kept in parallel until its last save in June 2022; in July 2022 a memo emailed to the revenue team retires it
Format change; what was lost or assumed	COMPLETENESS OWNERSHIP System import of public rates. **The import carries public rates, not the CORP tab or its conditions (the festival blackout is lost a second time); the workbook is not archived, relabeled or removed; the memo is an email, so no file anyone can search says the workbook is retired; separately, in 2023 the site owner leaves and nobody is assigned**
Accountable owner	Head of revenue (the system); nobody for the workbook after 2023
Who can change it	The revenue team (edit) and site collection administrators; nobody is responsible for it
Classification; retention or erasure duty	Personal data and commercial terms, retired by memo only
Read by (people, apps, models, agents)	Anyone in All staff (internal)
Cost per month; basis	Per month: the system subscription; the workbook costs nothing until someone reads it (illustrative)

Hop 7 is on the causal path. In March 2022 the group imported public rates from Rates master.xlsx into a rented revenue management system and maintained them there from then on; for the 2026 festival dates the harbor hotel's public rate is 236.00 euros. The workbook ran in parallel until its last save in June 2022, and in July 2022 a memo emailed to the revenue team retired it. The import carried public rates only, not the CORP tab or its conditions, so the festival blackout, already lost once in 1987, was lost a second time. The workbook was not archived, relabeled or removed, and because the memo was an email, no file anyone could search said it was retired. In 2023 the site owner left and nobody was assigned. From then on, All staff (internal) could still read the workbook and nobody was responsible for it.

THE HOP · 8 · 2024 · CORP TAB TO A CONSUMER CHAT ASSISTANT ON A PERSONAL ACCOUNT

Record	The rate sheet (hotel and retail group), hop 8, 2024
From	CORP tab
To	A consumer chat assistant on a personal account
Mechanism and schedule	A sales coordinator pastes the CORP tab into a chat to draft client emails
Format change; what was lost or assumed	RIGHTS Pasted text in a chat history. **Client contacts and rates sit in a chat history outside the group's retention and contracts; the failure is the group's missing rule on personal accounts**
Accountable owner	None
Who can change it	The coordinator
Classification; retention or erasure duty	Personal data and commercial terms outside the group's control
Read by (people, apps, models, agents)	The coordinator; the provider, as its terms allow
Cost per month; basis	Per month: nothing (a free account) (illustrative)

Hop 8 is a side branch. In 2024 a sales coordinator pasted the CORP tab into a consumer chat assistant on a personal account to draft client emails. Client contacts and rates now sat in a chat history outside the group's retention and contracts, readable by the coordinator and by the provider as its terms allow. The provider did what its service does. The failure is the group's: it had no rule on personal accounts for work data, no approved alternative and no way to see the paste. In March 2026 the incident team found this copy through an interview, not a tool, and asked the coordinator to delete the history.

ERA CARD

Reader. The incident team, the records manager and the auditor, who need a count of copies with its limits stated; and every holder of a copy, inside or outside the group.

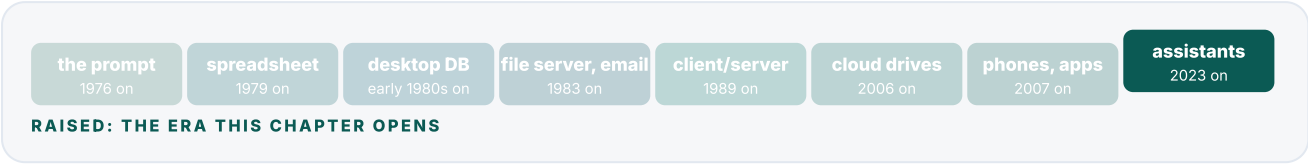
Security. Copies outside any list: imports, attachments, chat files, synced folders, hidden sheets, chat histories and backups, each with its own permissions and clock.

> **Cost.** Finding copies is cheap only before an incident; after one, each mailbox and export needs a decision and a record.
>
> **AI.** An assistant reads the copies a user may open, and its own histories are copies too; a personal chat account is a copy the organization cannot search.

Monday rule: pick the one record your pricing or revenue depends on, count its copies by location, sharing scope and authority state, write down which discovery methods you used and what they could not see, and set a rule for work data in personal chat accounts.

CHAPTER THIRTEEN

The Assistant Reads Your Files



On Tuesday 10 March 2026 a retail B2B account manager in the hotel and retail group's 300-seat assistant pilot asked for the best rate at the harbor hotel for a client's festival weekend. The assistant answered 189.00 euros, named two corporate clients and their contacts as holders of that rate, and cited Rates master.xlsx in the Revenue library. The rate was the corporate rate as last maintained in June 2022, and by conditions written nowhere in the workbook it did not apply on festival dates. The applicable rate, 236.00 euros, sat in a system the assistant could not see. The account manager emailed the answer to the client. Every step up to the answer was allowed: the account manager could open the library, the library was indexed, the assistant searched what the account manager could open, and bypassed no permission. This chapter explains how a workplace assistant chooses what to read, why it read this file and not the system that replaced it, and what an organization checks before it switches one on. The decision it changes: **before an assistant reaches a site's content, make the authoritative source reachable, the retired file unreachable and the broad groups narrow**.

Grounded in what you may open

A workplace assistant of this kind answers by retrieving content and passing it to a language model with the user's question. Microsoft announced that Microsoft 365 Copilot was "generally available for enterprise customers worldwide" on 1 November 2023. Its documentation states the rule that decides what it retrieves: "Microsoft Copilot only surfaces organizational data to which individual users have at least view permissions. It's important that you're using the permission models available in Microsoft 365 services, such as SharePoint, to help ensure the right users or groups have the right access to the right content within your organization." (Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot", page dated 9 July 2026, read 20 November 2026.) It adds that the semantic index "honors the user identity-based access boundary so that the grounding process only accesses content that the current user is authorized to access".

Google states the equivalent for Gemini in Google Workspace: "Gemini has the same access to Workspace data as you do", and it "can only access files in Drive that have been shared with you" (Google Workspace Learning Center, "What controls Gemini's access to Workspace data", read 20 November 2026). Google adds that if a Drive file's owner doesn't allow you to download, copy and print the file, "Gemini can't access it even if it's shared with you".

Both vendors describe the same design: the assistant inherits the user's reach. It is a consistent design, and it moves the whole question to the permissions. If a person may open a file, the assistant may read it for them, quote it and cite it; the controls listed below can narrow that. What used to protect an old file was obscurity: nobody knew where it was. Enterprise search weakened that protection; the assistant goes further, because it reads the file and writes a confident sentence from it.

Oversharing is inherited permission

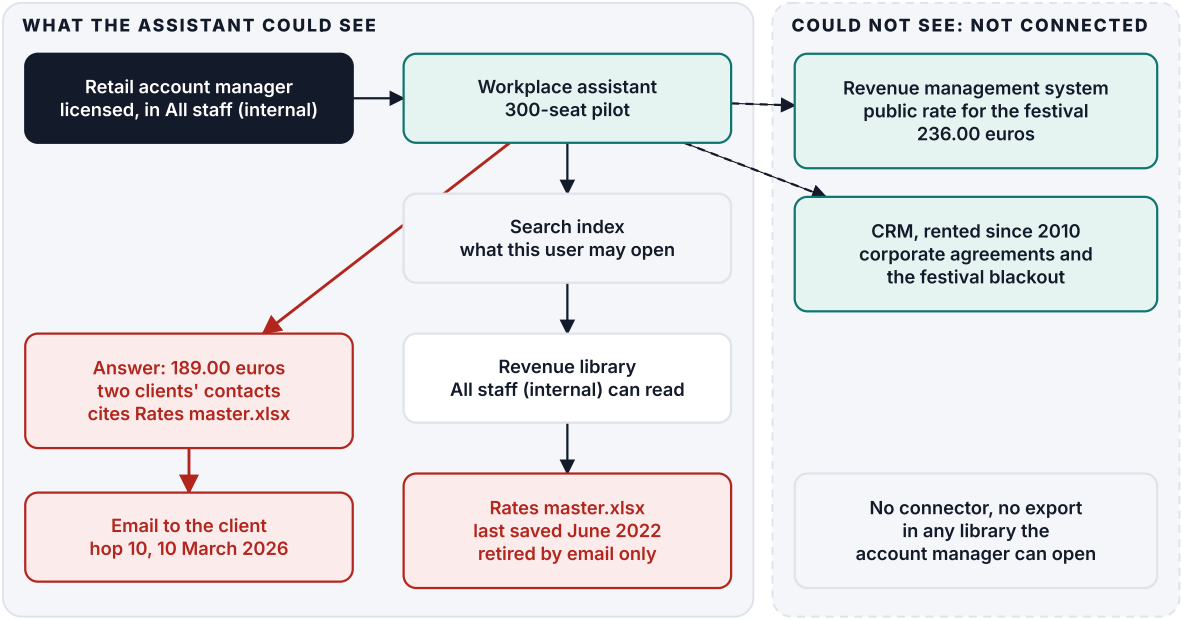
Oversharing, in this book, means a grant wider than the people who need the content. In the thread it took three decisions. In 2016 a migration mapped unresolved groups to one custom group, All staff (internal), and gave it read access to the Revenue library (hop 5). In 2018 an HR feed made every employee a member, growing the group to about 9,000, retail account managers included (hop 5b). In 2025 the pilot's readiness review looked for organization-wide links and the built-in everyone groups, and the custom group passed (hop 9). No one widened access on purpose after 2016. The grant was inherited by the library, then by every new member, then by the assistant.

The assistant also could not see the system that replaced the file. The revenue management system held the 236.00 euro public rate; the rented CRM held the corporate agreements, each with a festival blackout clause. Neither was connected to the assistant, and no export of either sat in any library the account manager could open. Microsoft documents that data from Graph connectors can be returned "if the user has permission to access that information", so a connection was possible; the group had not built one. With no better source in reach, the assistant took the one file that answered the question in words and numbers. Nothing in that file said it was retired: the retirement was an emailed memo of July 2022.

Figure 13.1 draws what the assistant could and could not see. It shows the Prologue's three failures: authority (hop 7), a file retired only by memo; access (hops 5 and 5b), a library open to about 9,000 people; and retrieval (hop 9), an assistant that could retrieve the workbook and could not reach the revenue management system or the CRM. Each is a defect on its own; fixing one leaves the other two.

FIGURE 13.1

What the assistant could and could not see: the retired workbook through a broad group, not the systems that replaced it



The account manager's permissions reached the Revenue library through All staff (internal), so the assistant could read the retired workbook; the revenue management system and the CRM held the right answer and were not connected.

Author's diagram based on the book's running example and Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" (read 20 November 2026); the hotel and retail group is illustrative.

Features that limit retrieval, as documented

Microsoft documents several features that narrow what Microsoft 365 Copilot reads. They work in different ways, and none of them is a substitute for correct permissions. The list below follows Microsoft Learn as read on 20 November 2026; it does not rank them.

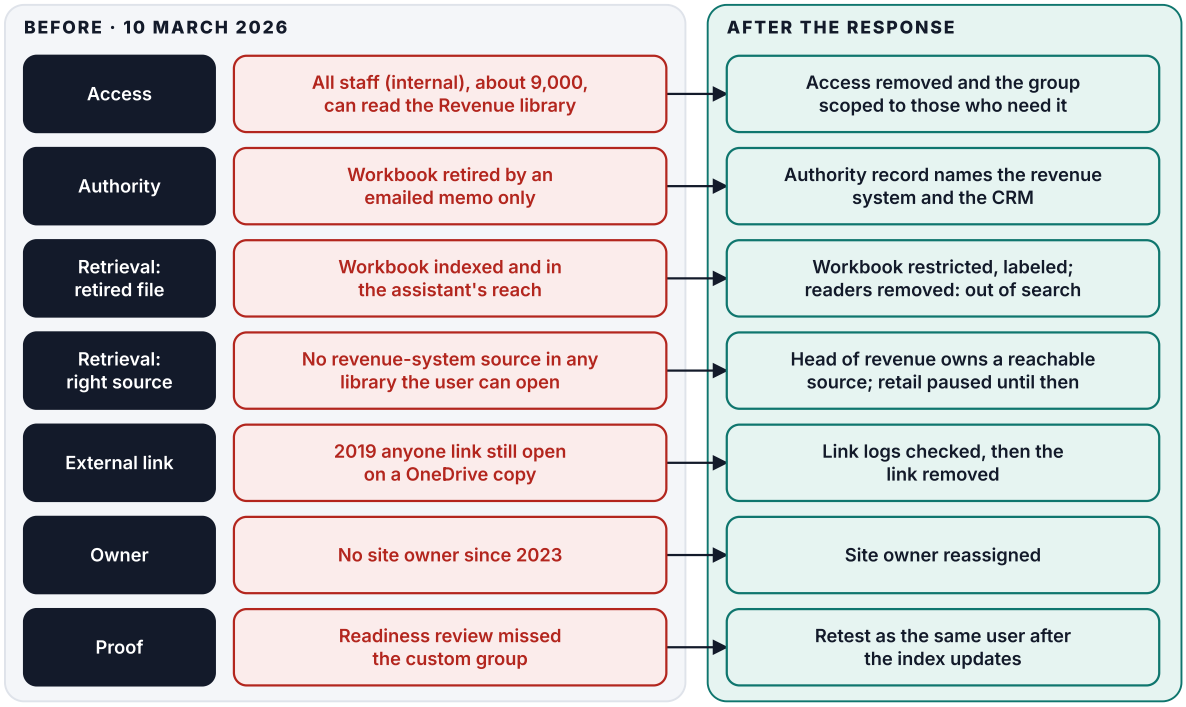
- **Sensitivity labels with encryption.** Microsoft documents that Copilot honors the usage rights of encrypted content, and that the EXTRACT right is the one that "determines whether Copilot or agents can display text to the user from encrypted content". A label that applies no encryption does not by itself keep content out of answers.
- **Data loss prevention for the Copilot location.** A policy can "prevent these Copilot experiences from using files and emails that have sensitivity labels when generating responses". Microsoft adds that such items "still appear in the citations of the response". The label marks the file; the policy acts on the mark.

- **Restricted Content Discovery.** A site-level setting that limits a site's content in organization-wide search and Copilot "while those reviews are taking place". Microsoft states that it "doesn't change existing permissions", does not apply to OneDrive, and "doesn't remove content from the Microsoft 365 search index"; on large sites it can take more than a week to take effect.
- **Site access reviews.** Part of SharePoint Advanced Management: an administrator sends a data access governance report, such as the "oversharing baseline report using permissions", to a site's owner, who reviews and removes access. Microsoft notes they do not support OneDrive, and they need an owner to answer: the Revenue site had none after 2023.
- **Restricted SharePoint Search.** An allow list of up to 100 sites for organization-wide search and Copilot. Microsoft states it "is retiring. Starting July 31, 2026, new enablement is blocked", and points to Restricted Content Discovery instead.

These features change what the assistant shows, not who can open the file. Only removing the grant, or moving a retired file to an archive location the user cannot open, takes it out of the user's reach and so out of the assistant's. Chapter Fifteen's fix does both and retests as the same user; Figure 13.2 sets each fix against its gap.

FIGURE 13.2

Before and after the response: the three failures (access, authority and retrieval, in two rows), then the link, the owner and the proof



The response narrows access, records authority, takes the retired workbook out of search, gives the decision on a reachable source an owner while retail use stays paused, removes the external link, names a site owner and proves the result by retesting as the same user.

Author's diagram based on the book's running example (the incident response told in Chapter Fifteen); illustrative.

The assistant as writer

An assistant also makes copies. Microsoft documents that it stores "the user's prompt and Copilot's response, including citations", the activity history of Chapter Eleven. A draft written in Word or Outlook is a new file or message the moment it is saved or sent. Microsoft documents that Copilot Pages are stored in a "user-owned SharePoint Embedded container", private by default, so a page built from an answer is one more copy in a personal store. Hop 10 is the plain case: the account manager pasted the answer into an email, which made a sent item inside the group and a recipient-held copy outside it. Every copy an assistant helps make goes into the census of Chapter Twelve, with its own owner, scope and deletion reach.

The assistant readiness check

Ten questions, each tied to the hop where the rate sheet failed it. Ask them before a pilot widens, for each important record in reach; Appendix A carries the same check.

#	Question	Hop
1	For the decisions these users ask about most, is the authoritative source named in an authority record?	7
2	Can the assistant reach that source, through a connector or a dated export, under these users' permissions?	9
3	Are replaced files retired four ways (location, permissions, label and a register entry), not by memo?	7
4	Has every group that grants read on content in the index been reviewed by name and member count, custom and HR-fed groups included, not only the built-in everyone groups?	5, 5b, 9
5	Does every site and library in the index have a current owner?	7
6	Are anyone links and personal-drive copies of governed files listed, and expired or removed on sites that hold personal data?	6
7	Do files in the index hold personal data their readers have no purpose to see, such as a pasted contact or guest tab?	4

#	Question	Hop
8	Do tabs that look like live extracts still refresh from their source, and do they show their owner and last refresh?	4b
9	Can an erasure be traced to every copy the assistant can read: versions, drives, mail, histories?	6b
10	Is there a rule for where answers and work data may go (client email, personal chat accounts), and are prompts, answers and citations kept long enough to trace an answer to its copy and state?	8, 10

This book covers which copies exist and how far a deletion reaches; how an agent is granted access at runtime belongs to the companion Medallion series. The checks above decide what an assistant can find.

THE HOP · 9 · 2025 · REVENUE LIBRARY AND OTHER FILES TO WORKPLACE ASSISTANT (300-SEAT PILOT)

Record	The rate sheet (hotel and retail group), hop 9, 2025
From	Revenue library and other files
To	Workplace assistant (300-seat pilot)
Mechanism and schedule	The assistant grounds answers through the service's search over content each user is permitted to access
Format change; what was lost or assumed	CONTEXT RIGHTS Search index over SharePoint and OneDrive; answers cite their source files; the CORP tab is the workbook's first sheet and holds a small table. **The pilot is still active in March 2026; the account manager holds a licence and is in All staff (internal); the current library item (last saved June 2022; opens before the audit window cannot be shown) inherits the library's permissions and is indexed; nothing marks it as retired; no revenue-system export sits in any library the account manager can open; the revenue management system and the CRM are not connected; the pilot's readiness review looked for organization-wide links and the built-in everyone groups, and the custom group passed**
Accountable owner	Head of digital workplace (the assistant); nobody for the workbook
Who can change it	Assistant administrators; tenant search settings
Classification; retention or erasure duty	Personal data and commercial terms, as in the files it reads; the assistant inherits every permission the user has
Read by (people, apps, models, agents)	300 pilot users, including the retail division's B2B account managers
Cost per month; basis	Per month: pilot licences for 300 seats (illustrative)

Hop 9, seen from the system side. The pilot started in 2025 and was still active in March 2026. The account manager held a licence and was a member of All staff (internal). The current library item, Rates master.xlsx, last saved in June 2022, inherited the library's permissions and was indexed. Its first sheet was the CORP tab. Nothing marked the file as retired, no revenue-system export sat in any library the account manager could open, and the revenue management system and the CRM were not connected. The assistant grounded its answer through the service's search over content the user was permitted to access. CONTEXT was lost because the conditions of the corporate rate, the festival blackout among them, were not in the file; RIGHTS because client contacts that only the sales team could read in the CRM were readable by a retail account manager in 2026.

THE HOP · 10 · 2026 · ASSISTANT ANSWER TO EMAIL TO A CLIENT (A RECIPIENT-HELD COPY OUTSIDE THE GROUP, AND A SENT ITEM INSIDE)

Record	The rate sheet (hotel and retail group), hop 10, 2026
From	Assistant answer
To	Email to a client (a recipient-held copy outside the group, and a sent item inside)
Mechanism and schedule	The account manager pastes the answer into an email on Tuesday 10 March 2026
Format change; what was lost or assumed	TIME RIGHTS Email text with a rate and two clients' names and contacts. **A corporate rate nearly four years old that does not apply on festival dates (189.00 euros; the applicable rate is the public 236.00) and two clients' names and contacts leave the group**
Accountable owner	Retail division sales
Who can change it	The account manager
Classification; retention or erasure duty	Personal data disclosed to a third party; commercial terms
Read by (people, apps, models, agents)	The client and anyone they forward it to
Cost per month; basis	One-off: lost margin on one quote and a breach assessment (not monthly) (illustrative)

Hop 10 is the step that turned an answer into a disclosure. On Tuesday 10 March 2026 the account manager pasted the answer into an email to the client: a corporate rate nearly four years old that does not apply on festival dates, 189.00 euros against an applicable 236.00, and the names and contacts of two corporate clients. A copy now sat in the client's inbox, outside the group, and a sent item inside it. On Thursday 12 March the client replied quoting 189.00 and asking for the same terms as one of the named companies, and the revenue manager traced the answer. Chapter Fifteen tells the response.

ERA CARD

Reader. The workplace assistant, reading for each user everything that user may open, and the user who trusts the citation.

Security. Inherited permissions decide the reach; labels, data loss prevention, discovery limits and access reviews narrow it.

> **Cost.** Licences are the visible cost; the larger ones are the clean-up of broad groups and retired files, and the rework when an answer leaves the building.
>
> **AI.** The assistant cannot reliably tell a retired file from a current one unless something it can read says so, and it cannot use a source it cannot reach.

Monday rule: run the ten readiness questions on one record your pilot users ask about every week, and do not widen the pilot until the authoritative source is reachable and the retired copies are not (questions 2 and 3).

Part V ends with the whole thread in one table: every hop of the rate sheet from 1987 to 2026, in the Ledger's ten columns, including the response (hop 10b) that Chapter Fifteen tells.

THE LEDGER, PART 1 OF 2

One Sheet, Since 1987: the full Ledger: movement and meaning

Hop, year	From	To	Mechanism and schedule	Format change; what was lost or assumed
0 1987	Reservations manager's notes and the typed rate memo	RATES.WK1 on a floppy disk (Lotus 1-2-3 Release 2, MS-DOS PC)	The reservations manager types the season's rates into a worksheet, prints it and faxes the printout to 12 hotels	CONTEXT Worksheet cells: hotel, room type 01 to 12, date range, rate code (RACK, CORP), rate; the seasonal uplift computed by a 1-2-3 macro. **The conditions of each rate (cancellation, deposit, who qualifies for the corporate rate, and blackout dates when corporate rates do not apply, such as the harbor festival) stay in the typed memo; the sheet holds codes and numbers**
1 1994	RATES.WK1	RATES94.XLS on a NetWare file server (S:\REVENUE)	The sheet is opened in Excel 5 on Windows for Workgroups 3.11, rebuilt and saved as an Excel workbook on the head office file server	MEANING Excel 5 workbook (BIFF5); NetWare trustee rights on the S:\REVENUE directory. **Excel could run the 1-2-3 Release 2 uplift macro while the file stayed WK1, but nobody translated it into an Excel macro (Excel 5 offered a translation assistant; the team did not use it); the rebuilt Excel workbook does not contain the macro, staff paste the values, and next season's rates can no longer be derived from the base rate**
1b 1996	The harbor hotel's copy of RATES94.XLS (sent each season on a floppy disk)	The harbor hotel's Access group-bookings database (linked table)	The harbor hotel's sales coordinator links the hotel's local copy of the workbook into an Access 95 database that prices group bookings	TIME Access .mdb with a linked table pointing at C:\RATES\RATES94.XLS on the hotel PC. **When the sheet arrives by email in 1998 under a new name, as an Excel 97 workbook, a version not in the list Access 95's help gives for linking, and last season's file is cleared from C:\RATES, the link fails on open and nobody relinks it; staff import last season's sheet from its floppy disk as a local copy and quote group rates from it for two seasons**

Hop, year	From	To	Mechanism and schedule	Format change; what was lost or assumed
2 1998	Head office file server	Email attachment to 40 hotels, one copy per hotel	Each season the sheet is emailed to every hotel; general managers save and edit their own copies; the practice continues until the 2022 replacement	OWNERSHIP Excel 97 workbook (hotels on older Excel cannot open it until they upgrade); long file names such as RATES98 final GM.xls. **Each hotel saves and edits its own copy; head office's S: copy stays the master by habit, not by any record**
3 2003	Head office workbook	Rented booking engine (extranet upload), one CSV per hotel	Head office exports one CSV per hotel; each hotel's front office manager edits it and uploads it to the booking engine's extranet; web rate codes are created (the ancestors of today's BAR codes)	IDENTITY TIME MEANING PRECISION CSV text; EU hotels save semicolon-delimited files with day-first dates and decimal commas under their regional settings; the engine parses with US settings. **Room types 01 to 09 lose their leading zeros when a hotel opens and resaves its CSV in a spreadsheet; 04/05 is read as 5 April, so up to 132 dates a year swap silently; in the semicolon-delimited files the comma in 189,00 is read as a thousands separator, so two hotels publish 18900 until a guest complains; the engine's rate field holds whole units, so 189.50 becomes 190. The engine reads dates month first and falls back to day first when the month would exceed 12, so only ambiguous dates swap; most hotels caught the 18900 rates on checking their prices, two did not; 189.50 comes from files written with a decimal point**
4 2004	On-premises CRM on SQL Server	The CORP tab, the workbook's first sheet (ODBC query, refreshed on open), and a VIP tab (pasted snapshot)	An analyst adds an ODBC query that pulls corporate accounts, negotiated rates and contacts into a CORP tab, and pastes the top loyalty guests into a VIP tab	RIGHTS Excel query table over ODBC; contact names, business emails and phone numbers; guest names and preferences. **Client contacts and top guests, readable in the CRM only by the sales team, become readable by everyone with rights to S:\REVENUE and by every hotel that receives the sheet; no recorded purpose, owner or retention period (the EU hotels then under national laws implementing Directive 95/46/EC)**
4b 2010	RATES94.XLS and its descendants	Rates master.xlsx (Excel 2010)	The workbook is saved in the new format; the CRM moves to a rented service and the on-premises database is switched off	TIME Office Open XML workbook (.xlsx); the CORP query's ODBC source no longer exists. **The CORP query fails on refresh; staff dismiss the warning, then remove the connection and maintain the tab by hand (contacts frozen at 2010; rates typed by hand) until the last save in June 2022; to readers outside the revenue team it still looks like a live extract. Corporate agreements of record move to the rented CRM, which is never connected to the workbook or, later, to the assistant**
5 2016	Windows file server (groups remapped from NetWare trustees in 2006)	SharePoint Online library Revenue	A third-party migration tool moves the share to the cloud; to finish on time the team maps unresolved and nested groups to one custom group	RIGHTS Files start a version history in the library; permissions inherited from the library; no item keeps narrower permissions. **The custom group All staff (internal), then the hotel and head office staff (about 4,000 people), gets read access to the Revenue library; the corporate tab becomes readable by all of them**

Hop, year	From	To	Mechanism and schedule	Format change; what was lost or assumed
5b 2018	HR system	The group All staff (internal)	When the retail division moves into the same tenant, an HR feed makes every employee a member of All staff (internal); nobody reviews what the group can reach	RIGHTS Dynamic group membership from the HR feed. **The group grows from hotel and head office staff to about 9,000 people, including the retail division's account managers; the Revenue library's read grant grows with it**
6 2019	Revenue library	A general manager's OneDrive copy, shared with an events agency by an anyone-with-the-link link	A general manager saves a copy to her OneDrive and sends the agency a link that needs no sign-in	OWNERSHIP RIGHTS rates final GM v2.xlsx; a link with no expiry. **A copy outside the Revenue library's permissions; anyone holding the link can open it; the general manager, still employed, moves to another hotel in 2021 and the copy is never reviewed; found only in 2026**
6b 2020	Erasure request from an EU guest (GDPR Art. 17)	The CRM and the library's current file	The privacy team removes the guest from the CRM and deletes the row from the current workbook	RIGHTS Deletion in two places. **The erasure duty does not travel: the guest stays in the library's version history, the 2019 OneDrive copy, mailbox attachments since 2004 and backups**
7 2022	Rates master.xlsx	Rented revenue management system (one import)	Public rates are imported in March 2022 and then maintained in the new system (236.00 euros: the harbor hotel's public rate for the 2026 festival dates); the workbook is kept in parallel until its last save in June 2022; in July 2022 a memo emailed to the revenue team retires it	COMPLETENESS OWNERSHIP System import of public rates. **The import carries public rates, not the CORP tab or its conditions (the festival blackout is lost a second time); the workbook is not archived, relabeled or removed; the memo is an email, so no file anyone can search says the workbook is retired; separately, in 2023 the site owner leaves and nobody is assigned**
8 2024	CORP tab	A consumer chat assistant on a personal account	A sales coordinator pastes the CORP tab into a chat to draft client emails	RIGHTS Pasted text in a chat history. **Client contacts and rates sit in a chat history outside the group's retention and contracts; the failure is the group's missing rule on personal accounts**

Hop, year	From	To	Mechanism and schedule	Format change; what was lost or assumed
9 2025	Revenue library and other files	Workplace assistant (300-seat pilot)	The assistant grounds answers through the service's search over content each user is permitted to access	CONTEXT RIGHTS Search index over SharePoint and OneDrive; answers cite their source files; the CORP tab is the workbook's first sheet and holds a small table. **The pilot is still active in March 2026; the account manager holds a licence and is in All staff (internal); the current library item (last saved June 2022; opens before the audit window cannot be shown) inherits the library's permissions and is indexed; nothing marks it as retired; no revenue-system export sits in any library the account manager can open; the revenue management system and the CRM are not connected; the pilot's readiness review looked for organization-wide links and the built-in everyone groups, and the custom group passed**
10 2026	Assistant answer	Email to a client (a recipient-held copy outside the group, and a sent item inside)	The account manager pastes the answer into an email on Tuesday 10 March 2026	TIME RIGHTS Email text with a rate and two clients' names and contacts. **A corporate rate nearly four years old that does not apply on festival dates (189.00 euros; the applicable rate is the public 236.00) and two clients' names and contacts leave the group**
10b 2026	The incident	A recorded response, then retirement done properly	Awareness recorded at 11:20 on Thursday 12 March 2026 (the 72-hour point is 11:20 on Sunday 15 March); an incident lead works with the CISO, the data protection officer, legal, records and the revenue owner; evidence first, then containment, then clean-up	Incident record, evidence set, risk decisions, authority record, restricted archive. **No new loss: evidence preserved before any change; the workbook restricted, not edited; Art. 33 decisions documented by the group with the data protection officer (the email: no notification; the 2019 link: notified); the clients and the recipient handled; access, link and group fixed; site owner reassigned; workbook taken out of search by removing every reader but the incident team; the workbook kept unedited as an evidence copy under Art. 17(3)(e), open only to the incident team, with a review date; retrieval retested**

THE LEDGER, PART 2 OF 2

One Sheet, Since 1987: the full Ledger: control, readers and cost

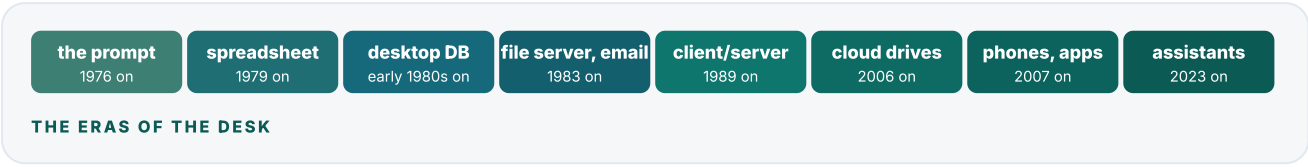
Hop	Accountable owner	Who can change it	Classification; retention or erasure	Read by	Cost per month; basis
0	Reservations manager, head office	The reservations manager (the only machine-readable copies are on her floppy disk and her PC; the 12 hotels hold faxed paper)	Commercial pricing data; no retention rule	Reservations staff; 12 hotels reading the faxed printout	Per month: one PC and a software licence; per season: about two days of work (basis: staff time) (illustrative)
1	Reservations manager, head office	Anyone with trustee rights to S:\REVENUE (the reservations team, two IT staff)	Commercial pricing data; no retention rule	Head office reservations and finance; each hotel from a floppy copy each season (1994 to 1998)	Per month: a share of the file server; no change in staff time (illustrative)

Hop	Accountable owner	Who can change it	Classification; retention or erasure	Read by	Cost per month; basis
1b	None recorded (built by the harbor hotel's sales coordinator)	The harbor hotel's sales staff	Commercial pricing data; no retention rule	Group sales staff at one hotel	Per month: near zero; one-off: two seasons of mispriced group quotes (not monthly) (illustrative)
2	None recorded for the hotel copies; the reservations manager for the S: copy	Every general manager, and anyone a copy is forwarded to	Commercial pricing data; no retention rule	40 general managers and their front office staff	Per month: mail storage near zero; per season: about three days of reconciliation (basis: staff time) (illustrative)
3	Head of distribution (engine contract); front office managers (uploads)	Each hotel's front office manager; the engine vendor's support staff	Commercial pricing data published to the web	Guests on the web, the booking engine, channel partners	Per month: the engine fee per booking; one-off: rework of the mispriced dates (not monthly) (illustrative)
4	None recorded for the tabs	The analyst who wrote the query; anyone who can save the workbook	Personal data of business contacts and guests; commercial terms	Head office, 40 hotels, anyone a copy reaches (the hotels read the rows as of the last save at head office; refresh ran only on the head office network; at the hotels each opening showed a failed refresh)	One-off: an ODBC driver and about 4 hours of work, an afternoon (not monthly) (illustrative)
4b	Head of revenue (a new role, 2009)	The revenue team; anyone with rights to the share	Personal data and commercial terms; no retention rule	The revenue team; hotels receiving the seasonal email; anyone with rights to the share	Per month: a few hours of manual upkeep (basis: staff time) (illustrative)
5	Head of revenue (site owner)	The revenue team (edit); site collection administrators	Personal data and commercial terms; no retention policy set on the library	All staff (internal) (read): about 4,000 staff in 2016, few of whom need the CORP tab	One-off: the migration project (not monthly); per month: a share of licences (illustrative)
5b	Head of IT operations (identity)	The HR feed; identity administrators	Group membership; no access review of what the group can reach	Every member of the group, and every system that trusts it	Per month: near zero; the cost is the reach (illustrative)
6	The general manager in name (personal drive); not reviewed after her 2021 move	The general manager; anyone with the link can view	Personal data and commercial terms outside the library's controls	The events agency and anyone the link was forwarded to (unknown)	Per month: near zero (illustrative)

Hop	Accountable owner	Who can change it	Classification; retention or erasure	Read by	Cost per month; basis
6b	Privacy team	The privacy team; the revenue team	Personal data under an erasure request	Anyone who can open an old version or a copy	One-off: 2 hours of work (not monthly); the exposure stays open (illustrative)
7	Head of revenue (the system); nobody for the workbook after 2023	The revenue team (edit) and site collection administrators; nobody is responsible for it	Personal data and commercial terms, retired by memo only	Anyone in All staff (internal)	Per month: the system subscription; the workbook costs nothing until someone reads it (illustrative)
8	None	The coordinator	Personal data and commercial terms outside the group's control	The coordinator; the provider, as its terms allow	Per month: nothing (a free account) (illustrative)
9	Head of digital workplace (the assistant); nobody for the workbook	Assistant administrators; tenant search settings	Personal data and commercial terms, as in the files it reads; the assistant inherits every permission the user has	300 pilot users, including the retail division's B2B account managers	Per month: pilot licences for 300 seats (illustrative)
10	Retail division sales	The account manager	Personal data disclosed to a third party; commercial terms	The client and anyone they forward it to	One-off: lost margin on one quote and a breach assessment (not monthly) (illustrative)
10b	Head of revenue (authority record); the group as controller, advised by the data protection officer (assessments)	The named incident team only	Incident and assessment records; retention per policy	Auditors; the supervisory authority on request	One-off: about 160 hours across eight people (not monthly) (illustrative)

CHAPTER FOURTEEN

Authority and Inventory



On Thursday 12 March 2026 the revenue manager traced the answer the account manager had emailed two days before. Finding the file was the easy part: the assistant had cited Rates master.xlsx in the Revenue library, and its file properties showed a last save in June 2022. The hard question came next. Which source *should* have answered? The public rate for the festival dates was 236.00 euros in the revenue management system. The corporate rate and its festival blackout lived in each corporate agreement in the CRM. Nobody had ever written either fact down as a rule, so nothing a search index could read said that the workbook was no longer the answer. This chapter covers days 1 to 30 of a 90-day plan, and its first decision is the one the group never made: **name the authoritative source for each decision before you count, narrow or delete anything**.

Authority first: one source per decision, written down

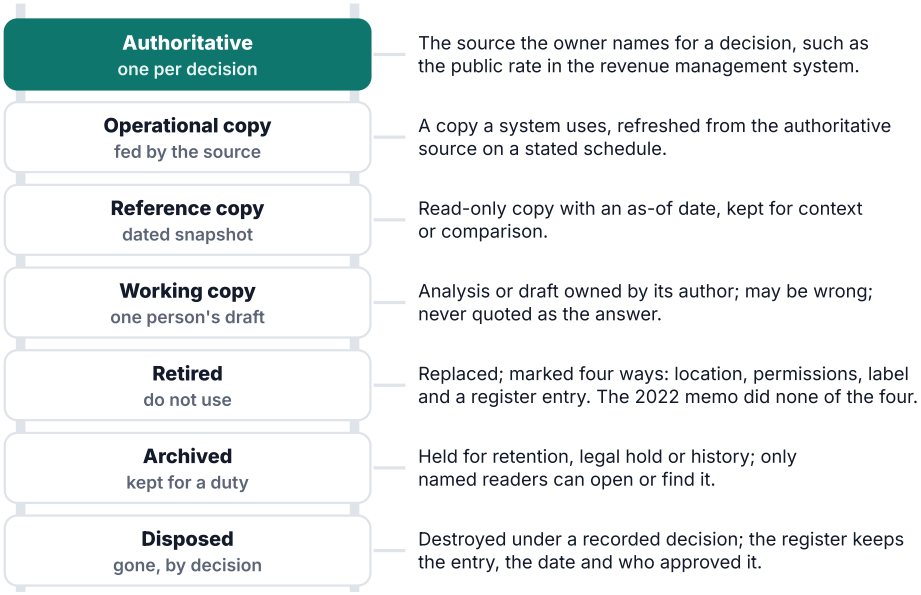
The Ledger tells you which copy an answer used and what state that copy was in. It cannot tell you which copy the answer should have used: a perfect trace from the assistant back to RATES.WK1 in 1987 is still a trace to the wrong file. That needs an **authority record**: for one business decision, the source the owner names as the answer, the owner's name, the date, and the state of every other copy that touches the decision.

Chapter Two introduced the seven authority states. Figure 14.1 puts them in order. Only the top rung decides an answer; an operational copy may serve it within its stated refresh window. When its job ends, a copy moves toward retired, archived and disposed, not always one rung at a time, and each move is recorded. Two rules matter most. First, authority belongs to a decision, not to a file: the revenue management system is authoritative for the public rate and has nothing to say about a corporate blackout. Second, **retired** is a state of its own, meaning "do not use for decisions", and it must be readable by people and machines

alike. Chapter Fifteen shows what that takes: a location, permissions, a label and an entry in a register. The group's July 2022 memo was an email to the revenue team. It retired the workbook in the minds of the revenue team and in no place an index could see.

FIGURE 14.1

Seven authority states: only the top rung decides an answer, and retirement is a rung of its own



Every copy of an important record sits on one rung for each decision it touches; when its job ends it moves toward retired, archived and disposed, not always one rung at a time, and each move is recorded in the authority register, not in a memo.

Author's model; the states are introduced in Chapter Two of this book.

Start with the decisions, not the files. Ask each business owner which decisions their team makes from data that lives on a desk or in a library, and for each one, where the answer should come from. For the rate sheet, the first authority record looks like this:

Decision	Authoritative source	Owner	Other copies and their state
Public rate for a hotel, room type and date	Revenue management system	Head of revenue	Rates master.xlsx: retired; hotel copies and the 2019 OneDrive copy: retired, then archived or disposed after the evidence is kept
Corporate rate and its conditions, including the festival blackout	The corporate agreement in the CRM	Head of sales (illustrative)	The CORP tab: retired; it has looked like a live extract since its query died in 2010 (contacts as they were in 2010, rates typed by hand until June 2022)
Which guests and contacts a revenue team may see	The CRM's access model	Head of sales, advised by the data protection officer	The VIP tab: no recorded purpose for a working copy; dispose after any duty to keep it is recorded

The table makes the pattern visible. The festival blackout was lost in 1987 (hop 0) and again in the 2022 import (hop 7) because nobody had named its source. The clause already sat in each corporate agreement in the CRM; writing "the corporate agreement in the CRM" in one cell is the first time in 39 years that anyone has named it as the source. The festival quote needs both rows: the revenue management system for the public rate, and the CRM agreement for whether a corporate rate applies on those dates. The table also names the sources behind the third failure of the Prologue, retrieval: the assistant could retrieve the workbook and could not reach either system in the second column, because neither was connected to it, and no better source was in reach. An authority record does not fix retrieval, but it tells you which connection or which export would.

> *Lineage shows which copy was used. The authority record says which copy should have been. Write the second before you trust the first.*

Discovery and the desk file inventory

With the decisions named, find the files. Chapter Twelve set out the discovery methods and what each one cannot see: search and content inventories, audit logs, eDiscovery, SaaS inventories, data loss prevention reports, endpoint and sync telemetry, data security posture tools and interviews. Use that table to choose; do not expect one method to find everything. The group's census in Chapter Twelve found 23 copies of the rate sheet, three of them outside the group, and recorded what its methods could not see. Your inventory should end the same way: a count, plus a sentence on its limits.

For each file you find, record the five optional file columns of the Ledger (Appendix A): **path or URL**, **copies known**, **sharing scope** (person, group, organization, anyone with the link), **authority state** and **last opened**. Then score it on two short scales, shown in Figure 14.2.

- **Impact**, 0 to 5: does the file feed a system of record, a regulator, a price, a model or an assistant? Count the yeses. "Feeds an assistant" means it is in an index an assistant searches, whether or not anyone meant it to be.
- **Exposure**, 0 to 4: is it shared beyond the team that owns it; does it hold personal data; does more than one copy exist; is the owner absent? Count the yeses.

FIGURE 14.2

Rank desk files by impact against exposure: the rate sheet lands in the top corner, and an anyone link to personal data is a candidate anywhere



Count the impact tests a file meets (it feeds a system of record, a regulator, a price, a model or an assistant) and the exposure tests (sharing beyond the team, personal data, more than one known copy, no owner present); the line is 2 on each scale, personal data behind an anyone link is a candidate whatever its impact, the rate sheet scores 2 of 5 and 4 of 4, and the other files are illustrative.

Author's diagram; the rate sheet from this book's running example (illustrative); other files invented for illustration.

Rates master.xlsx scores 2 for impact (a price, an assistant) and 4 for exposure: about 9,000 readers through All staff (internal), contacts and guest names in two tabs, 23 known copies, and no owner since the site owner left in 2023. A file with impact of 2 or more and exposure of 2 or more goes on the candidate list; so does any file with personal data behind an anyone link, whatever its impact score. Sort the list by exposure, then impact, then the number of readers, and take the first 20; exposure comes first because it is what the next 60 days can reduce. Twenty is not magic. It is how many files a small team can fix in the next 60 days, each with an owner and a Ledger row. If your list has 400 candidates, that is a finding to report, and the next 20 wait for the next quarter.

Rank	File (illustrative except row 1)	Impact	Exposure	Owner found	First action
1	Rates master.xlsx, Revenue library	2: price, assistant	4	None since 2023	Assign owner; mark retired; restrict
2	Loyalty export CSV in a OneDrive	2: a model, an assistant	3	Author, still employed	Move to the loyalty site; narrow sharing
3	Store pricing overrides workbook	3: system of record, price, assistant	2	Retail pricing lead	Authority record; Ledger row
Not ranked	Monthly revenue close workbook	2: system of record, regulator	1	Group financial controller	Below the line on exposure; owner and authority record only

The last row shows the rule working. The close workbook matters a great deal, but one owner keeps it inside one team, and its one exposure point is a second copy kept for the auditors, so it gets an authority record and stays off the top 20 for now. Ranking is how you answer the fair objection from Chapter Four: spreadsheets are how the business adapts, and banning them fails. You are not banning anything. You are choosing the files where a wrong copy reaches a price, a regulator or an assistant, and leaving the rest alone.

An owner for every file and every site, including after people leave

An owner is a person with a job title who can say yes or no to a change and who answers when the file is wrong. A team is not an owner, and neither is a platform. Every one of the top 20 gets an owner, and so does every site or library that holds one of them. If nobody will accept a file, write "none recorded" and the loss code OWNERSHIP; that is a finding, and it usually means the file should move down the ladder.

Ownership decays in three ways, and the thread shows all three. A person **leaves**: the Revenue site owner left in 2023 and nobody was assigned (hop 7). A person **moves**: the general manager who made the 2019 anyone link moved to another hotel in 2021, stayed

employed, and her copy was never reviewed (hop 6). A **group drifts**: an HR feed grew All staff (internal) from about 4,000 to about 9,000 people in 2018 and nobody reviewed what it could reach (hop 5b). Leaver processes catch only the first. Write a rule for each.

The platforms help with leavers, within limits you should read before relying on them. Microsoft documents, as of its OneDrive retention page dated 2 June 2026, that when a user is deleted the user's manager is given access to the user's OneDrive by default, that a secondary owner can be set for users without a manager, and that "if access delegation is disabled or a manager or secondary owner isn't set for a user, no one has automatic access". The default retention period is 30 days and can be changed. That 30-day clock starts only when the account is deleted: blocking sign-in or removing the licence does not start it. Removing the licence starts a different clock: Microsoft documents (page dated 26 August 2026) that unlicensed OneDrive accounts become read-only after 60 unlicensed days and are archived after 93. Google documents, on a page updated 10 November 2026, that an administrator can transfer a leaver's Drive files to another user, that items in Trash are not transferred, and that transferring files does not change who has access to them. That last sentence matters: a transfer gives a file a new owner, not a new audience, so an anyone link survives the move.

For sites, Microsoft documents a site ownership policy in SharePoint Advanced Management (page dated 14 May 2026) that finds sites below a minimum number of owners, can notify the managers of previous owners and recently active members, and can set an ownerless site to read-only or archive it after notices go unanswered. Microsoft also notes that a leaver's details are removed 30 days after they leave, so a manager may get only one notice. None of these tools knows that a general manager changed hotels. The mover rule is yours: when a person changes role, every file they share outside their new team is reviewed by their old manager within 30 days.

Event	Rule (illustrative wording)	Evidence
Leaver	Before the account is deleted, the manager accepts or reassigns each file and site the leaver owned; anyone links are removed	Leaver checklist signed; transfer log
Mover	Within 30 days, the old manager reviews files the mover shares outside the new team	Sharing report for the person, dated
Site with one owner	A second owner is named; ownerless sites go read-only after three notices	Site ownership policy report
Group change	Any group above 500 members (illustrative) lists what it can reach, reviewed quarterly	Access report for the group, signed by its owner

What it costs, and when to leave a file alone

A plan that does not state its cost will be cut at the first budget review. The table below is illustrative, for a group of this size and a top 20; replace every number with your own. It sets the cost of the work against the cost the group actually paid in March 2026.

Item	Basis (illustrative)	Estimate
Authority records	12 decisions, 3 hours each with the owner	36 hours
Inventory and scoring	Discovery runs plus interviews in 8 departments	80 hours
Owners and leaver rules	20 files and their sites, 2 hours each	40 hours
Clean-up of the top 20 (days 31 to 60)	6 hours per file: narrow, retire, Ledger row	120 hours
Governance tooling	Reports and policies at the tier you already license, or an added licence	Per vendor price list; often already paid
Storage	Archived copies in a colder tier	Small; storage is cheap, reconciliation is not
The rework of hop 10	The incident response in Chapter Fifteen: eight people	**About 160 hours**, plus a goodwill discount and a breach assessment
The pilot the clean-up protects	300 seats; Microsoft announced 30 dollars per user per month on 18 July 2023, before general availability on 1 November 2023	About 9,000 dollars a month at that list price

The comparison is not a forecast. It is a single incident against a single quarter of work. The four costed items of work together cost about 276 hours, once, against about 160 hours for one incident; the clean-up of the top 20 alone, about 120 hours, costs less than that one incident, and nothing in the group's estate made that incident the last. The hours for hop 10 are not the whole loss: the client relationship, the two named clients' trust and the separate assessment of the 2019 link sit outside any timesheet.

Leave it alone when any of these is true, and record the reason in one line:

- It feeds no system of record, regulator, price, model or assistant, and nobody outside its team reads it.
- It holds no personal data and no commercial terms, and it lives in one place with an owner present.
- It is a working copy its author will delete, inside a site that no assistant indexes.

- The decision it supports is reversible within a day and costs less than the hours to govern it.
- It is already archived under a recorded retention rule, readable only by the readers that rule names.

The list is the price of the plan's credibility. A team that tries to govern every desk file governs none of them, and the business stops answering its questions.

The first 30 days on one page

The work above fits in four blocks. Each ends with something the sponsoring executive can open and check, and each names who does it. The sponsor matters more than the tooling: the work crosses revenue, IT operations, privacy and the digital workplace team, and nobody below the executive committee owns all four.

Days	Output	Who	Done when
1 to 5	The decisions most asked of desk files and of the assistant pilot, with an authoritative source and owner for each	Chief data officer with business owners	Each owner has signed their row of the register
6 to 15	Discovery runs and interviews; every file found has the five file columns filled	IT operations, records, one analyst per department	The count is written with a sentence on what the methods could not see
16 to 22	Impact and exposure scores; the ranked list; the top 20	Chief data officer and CISO	The "leave it alone" reasons are recorded for the rest
23 to 30	An owner for each of the top 20 and their sites; leaver, mover and group rules approved	Business owners, HR, identity team	Every top 20 file has a name, or OWNERSHIP recorded as a finding

Two things are deliberately missing from the first 30 days: changing permissions and deleting files. Both come next, and both are safer once the authority record exists. A team that narrows access before it knows which source answers a decision will break the workflows that depend on the old copy, and will be told to put the access back. A team that

deletes before it counts will destroy the evidence it needs if the file has already caused harm, as the group learned in March 2026 when the first rule of its response was to change nothing until the evidence was kept.

ERA CARD

Reader. Every reader the rate sheet ever had, from the 1987 fax to the 2025 index; the authority record is written for the next one.

Security. Exposure is scored, not assumed: sharing scope, personal data, copies and an absent owner, file by file.

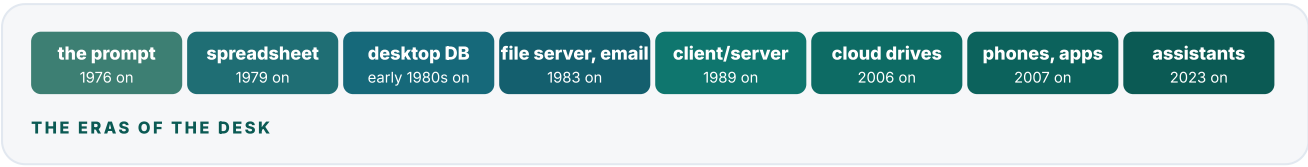
Cost. About 276 hours for authority, inventory, owners and the top 20 clean-up (illustrative), against about 160 hours to unwind one wrong answer.

AI. "Feeds an assistant" is an impact test: anything in the index is an input, whether or not anyone chose it.

Monday rule: pick the three decisions your assistant pilot is most often asked about, and write down the authoritative source and its owner for each before the week ends.

CHAPTER FIFTEEN

Permissions, Retirement and Proof

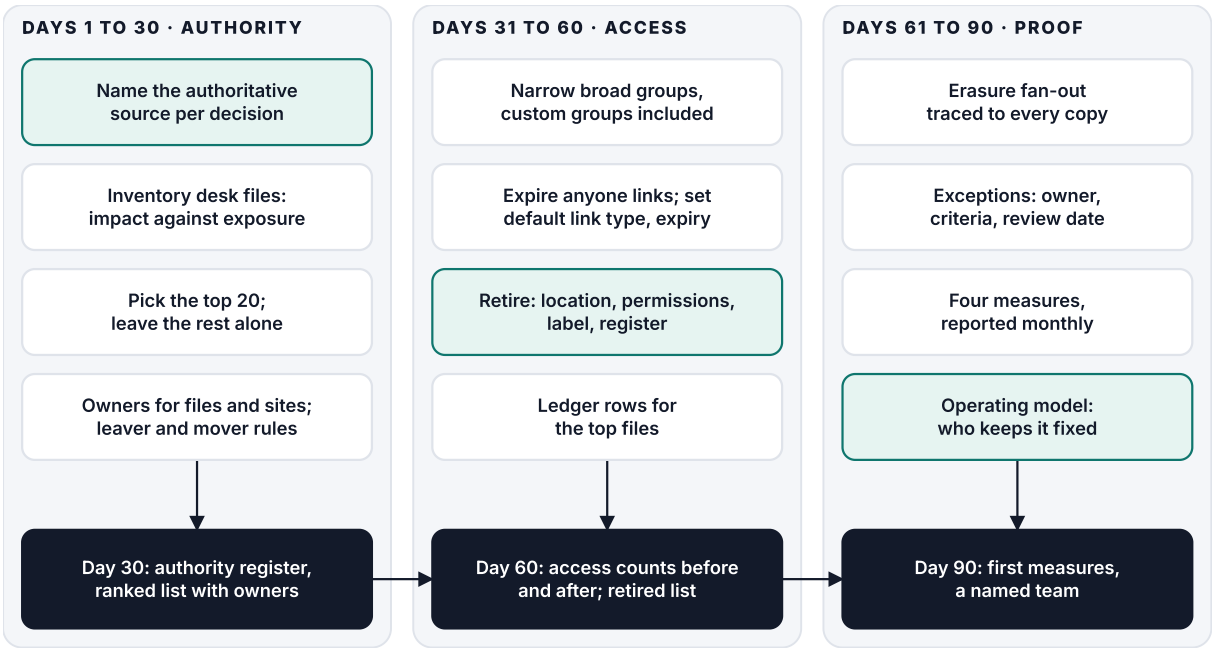


THE ERAS OF THE DESK

At 11:20 on Thursday 12 March 2026 the group recorded that it was aware of the incident. From that minute the first of two clocks ran: under Article 33 of the General Data Protection Regulation, a controller must notify a personal data breach to its supervisory authority "without undue delay and, where feasible, not later than 72 hours after having become aware of it", unless the breach "is unlikely to result in a risk to the rights and freedoms of natural persons". The 72-hour point was 11:20 on Sunday 15 March. The first instinct in the room was to open the workbook and fix it. The group did not, and that restraint is where this chapter starts. Days 31 to 90 of the plan narrow who can open the files, retire the replaced ones properly and build the proof; Figure 15.1 sets out the 90 days. The chapter ends by closing the thread: **evidence first, then containment, then clean-up, with every decision written down**.

FIGURE 15.1

The 90-day plan: authority first, access and retirement before each site enters the pilot, proof last



Each block of thirty days ends with something an executive can open and check; Chapter Fourteen covers the first block and this chapter the other two.

Author's diagram based on the plan in Chapters Fourteen and Fifteen.

Days 31 to 60: narrow the groups, then the links

Narrow broad groups, custom ones included. The pilot's readiness review looked for organization-wide links and the built-in everyone groups, and All staff (internal) passed, because it was a custom group (hop 9). The reports a team reaches for first are built for the built-in groups. Microsoft documents, on its data access governance page dated 9 July 2026, an activity report for content shared with "Everyone except external users" and a sharing links report, both covering the last 28 days, and snapshot reports, including "site permissions for users or groups", which lists every site a named user or group can reach. Use the last one on your own large groups. A group with about 9,000 members is an everyone group in all but name, and the report has to be asked about it by name.

For each of the top 20 from Chapter Fourteen, write down who can open it today, then decide who should. For the Revenue library the answer was the revenue team and the hotel revenue managers, not about 9,000 people. Remove the group's grant from the library rather than adding exceptions on top of it, and give the group itself an owner who must review, each quarter, what it can reach. Record the counts before and after; they are the day 60 evidence.

Expire anyone links and set the defaults. Microsoft documents, on its sharing settings page dated 30 June 2026, that "Anyone" links let people "access the files or folders without authenticating", that an administrator can require all Anyone links to expire within a set maximum number of days and can limit them to view only, and that when the expiry is shortened, existing links "update to the new setting", while a longer setting leaves existing links as they are. The same page lists the three link types an administrator can set as the default: Specific people, Only people in your organization, and Anyone with the link, the last available only when external sharing is set to Anyone. It warns that with Anyone links "you can't track who has access to shared items or who has accessed shared items". Microsoft's page on shareable links, dated 6 November 2026, adds the change Chapter Nine describes: one "hero link" per file, whose usual starting scope is "Only people added to the file have access", and narrowing that scope "applies to links already sent". Older links keep working, so they still have to be found and removed one by one.

The 2019 link to the events agency had no expiry. The settings exist; the rule did not. A reasonable default for a group like this one: the default link type is Specific people; Anyone links are off for sites that hold personal data, and elsewhere expire within 30 days and allow view only (illustrative numbers). Export the list of existing Anyone links before you shorten the maximum expiry. The shorter setting can expire old links, the page does not say how links created with no expiry are treated, and you need the list to review who was given what.

Retirement done properly: a memo is not retirement

A retired file is one nobody should use for a decision. To make that true for people and for machines, retirement has four parts, and each part is checkable:

Part	What it means	What the July 2022 memo did
Location	The file moves out of the working library into an archive location or tier	Nothing: the workbook stayed in Revenue/Rates
Permissions	Read access shrinks to the people with a duty to see it	Nothing: All staff (internal) could still read it
Label	A visible state ("retired, see the revenue management system") on the file or library	Nothing: the file looked current
Register	An entry in the authority register: the file, its new state, the replacement, the date, who approved	An email to the revenue team, which no index reads as a rule

The platforms offer tools for some of the four. Read their documentation for what each does not do. Microsoft documents (page dated 11 October 2026) that Restricted Content Discovery limits a SharePoint site's content in organization-wide search and Copilot responses, that it "doesn't change existing permissions", that it does not apply to OneDrive, that it does not remove content from the search index, and that for sites with more than 500,000 items an update can take more than a week. Microsoft describes it as "a temporary governance control" that gives organizations time to review access. Restricted SharePoint Search, an earlier allow-list control, is retiring: Microsoft states that "starting July 31, 2026, new enablement is blocked". Microsoft 365 Archive moves a site or file to a colder tier where, Microsoft documents (page dated 31 August 2026), content "is no longer directly accessible to anyone", yet the same page states that "Full content search works for Purview Content Search, end-user search, and eDiscovery search experiences". Archiving alone does not take a file out of search; removing who can read it does. None of these is a register. A label alone does not stop retrieval, and archiving is not deletion: Chapter Eleven drew that line, and it holds here.

Before or alongside the assistant. Days 31 to 60 do not have to finish before a pilot starts, but the order inside them matters. Narrow the groups and retire the replaced files for the sites the pilot will index first; widen the pilot only as each site passes the readiness check in Appendix A. The check's ten questions are tied to the hops where the rate sheet failed. Two of them are gates: do not widen the pilot to a site until the authoritative source is reachable (question 2) and the replaced files are out of reach (question 3). Every other "no" is a named risk with an owner and a date. The group's pilot review looked only for organization-wide links and the built-in everyone groups, a part of two questions. A short review that asks all ten is cheaper than the response that closes this chapter.

Ledger rows for the top files. Each of the top 20 gets its rows in the Ledger, with the optional file columns filled in and the authority state recorded (Appendix A shows a worked row for hop 7). Rows are written from the point of use back to the source, as the revenue manager walked back from the answer. A row with no loss is valid. A row with OWNERSHIP and "none recorded" is a finding for the owner rule in Chapter Fourteen.

Days 61 to 90: erasure fan-out, exceptions and measures

Trace every erasure to every copy. Article 17 gives a person the right to obtain "the erasure of personal data concerning him or her without undue delay" on the grounds it lists, and Article 19 requires the controller to communicate an erasure to "each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort". Copies the controller itself holds fall under Article 17 directly; Article 19 adds the recipients, such as the events agency. In 2020 the privacy team deleted an EU guest from the CRM and from the current workbook (hop 6b). The Ledger rows show

the other places the duty had to reach: the library's version history, the 2019 OneDrive copy, mailbox attachments since 2004 and backups. For each copy, write three things: the duty, whether this copy meets it, and how you would prove it. Where a copy cannot be cleared now, such as a backup that expires on its own schedule, record it as an exception with an end date. Do not claim more than you can show; the regulation does not say every backup is erased at once, and neither should your register.

Run an exceptions process. Every file kept outside the rules needs three things: an owner who accepts the risk, the criteria that justify the exception (a legal hold, a stated retention duty, a system that cannot yet export), and a review date no more than twelve months away (illustrative). An exception without a review date is a permanent policy nobody approved.

Measure four things monthly. Each measure is a count a board member could check.

Measure	Why it matters	Target (illustrative)
Top files with a named owner	OWNERSHIP was lost at hops 2, 6 and 7	20 of 20
Anyone links on sites with personal data, or older than 30 days elsewhere	The 2019 link lived about seven years unseen	Zero
Retired files still retrievable by an assistant user	Hop 9: the retired workbook was indexed and cited	Zero, tested by a search as a pilot user
Time to trace an answer to its copy and state	The proof the Ledger exists to give	Under one working day

Score the maturity. The scale is the author's, not a standard, and each level includes the one below.

Level	Name	What is true
1	Unknown	Copies and owners live in people's memory
2	Ranked	Authority records for key decisions; top 20 with owners
3	Narrowed	Broad groups scoped, link defaults set, retired files marked four ways
4	Proven	Erasures traced to every copy; exceptions dated; four measures monthly
5	Evidential	Any assistant answer traced to its copy and state on request, within a day

Name the operating model. The work stays fixed only if someone owns it after day 90. A workable split: the chief data officer owns the authority register and the measures; each business owner owns their decisions and files; IT operations owns groups, link settings and the leaver and mover rules; the privacy team owns erasure and exceptions that involve personal data, advised and monitored by the data protection officer; the head of the digital workplace owns what the assistant can index. One monthly meeting reviews the four measures and the exceptions due. If a measure goes the wrong way for two months, the owner explains why to the sponsor.

Closing hop 10b: the response, in order

The group's response to the March incident is the last hop of the thread, and it is a fix, not a loss. An incident lead worked with the CISO, the data protection officer (the role GDPR Article 37 requires in some organizations and Articles 38 and 39 set as an adviser and contact point), legal, records and the revenue owner. Article 4(12) defines a personal data breach as "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed". The team treated each exposure as a candidate breach and assessed each one separately.

THE HOP · 10B · 2026 · THE INCIDENT TO A RECORDED RESPONSE, THEN RETIREMENT DONE PROPERLY

Record	The rate sheet (hotel and retail group), hop 10b, 2026
From	The incident
To	A recorded response, then retirement done properly
Mechanism and schedule	Awareness recorded at 11:20 on Thursday 12 March 2026 (the 72-hour point is 11:20 on Sunday 15 March); an incident lead works with the CISO, the data protection officer, legal, records and the revenue owner; evidence first, then containment, then clean-up
Format change; what was lost or assumed	Incident record, evidence set, risk decisions, authority record, restricted archive. **No new loss: evidence preserved before any change; the workbook restricted, not edited; Art. 33 decisions documented by the group with the data protection officer (the email: no notification; the 2019 link: notified); the clients and the recipient handled; access, link and group fixed; site owner reassigned; workbook taken out of search by removing every reader but the incident team; the workbook kept unedited as an evidence copy under Art. 17(3)(e), open only to the incident team, with a review date; retrieval retested**
Accountable owner	Head of revenue (authority record); the group as controller, advised by the data protection officer (assessments)
Who can change it	The named incident team only
Classification; retention or erasure duty	Incident and assessment records; retention per policy
Read by (people, apps, models, agents)	Auditors; the supervisory authority on request
Cost per month; basis	One-off: about 160 hours across eight people (not monthly) (illustrative)

Evidence first. Before any change, the team preserved the prompt, the answer and its citation, the email, the audit and sharing logs, the permissions on the library and the item, the version history, and the assistant and search configuration. Microsoft documents that it stores a user's Copilot prompts and responses, "including citations to any information used to ground Copilot's response", and that administrators can find them with Content search or Microsoft Purview. The workbook was opened read-only during the trace, then restricted by

permissions and labeled at library level, not edited, so its last saved date stays June 2022. An edited file would have destroyed the plainest evidence that the answer was nearly four years stale.

The email: documented, not notified. Two business contacts, with no sensitive data, sent to one known recipient who was then asked to delete them, were assessed as unlikely to result in a risk, so the group, on the data protection officer's advice, decided not to notify the lead supervisory authority, and documented the decision with the data protection officer before 11:20 on Sunday 15 March. Article 33(5) is what makes that a decision rather than a silence: the controller "shall document any personal data breaches, comprising the facts relating to the personal data breach, its effects and the remedial action taken", and that documentation "shall enable the supervisory authority to verify compliance". Article 34 requires telling the people affected when a breach is "likely to result in a high risk"; it was considered and rejected. The recipient was asked to delete the email and confirm. The two named clients were told under their confidentiality terms. The 189.00 euro quote was not honoured; the client was offered the public rate with a goodwill discount.

The 2019 link: assessed separately, and notified. The anyone-link copy had its own awareness date and its own clock. The team checked the audit records for the link first, within the period they still covered: Microsoft documents an event logged "when an anonymous link is used to access a resource", which shows that the link was used and when, not by a named person. Then it removed the link and asked the agency to delete its copy. The copy held the VIP tab, including the guest whose erasure in 2020 never reached it, and it had been open without sign-in since 2019. The group notified the lead supervisory authority within 72 hours of that awareness, with the data protection officer as its contact point, and assessed telling the guests under Article 34.

The rest of the exposure. The internal over-access since 2016 was assessed separately, and it showed the limit of evidence: the audit logs did not reach back that far, so who opened the file could not be proven. Microsoft documents that its standard audit retention is 180 days for records generated from 17 October 2023 (90 days before that), with up to ten years available under added licences; a group that never set longer retention cannot answer a question about 2016. The 2024 paste of the CORP tab into a personal chat assistant was assessed separately as well, and the coordinator was asked to delete the chat history. The unfinished 2020 erasure was completed under Article 17. Her row was cleared from every other copy the group controls, current files and their version histories alike. Backups that cannot be edited were listed with their expiry dates. The workbook item, whose version history still held her row, was kept unedited as the evidence copy under the exception for legal claims (Article 17(3)(e)), open only to the incident team and given a review date.

Clean-up, defined. The team searched mailboxes, exports, version histories, OneDrive copies and the assistant index for the workbook and its personal data, swept everything the group could reach, and searched for other answers that had cited the file. It defined removal before doing it: current files and version history cleared, archives restricted, and only copies under legal hold or a stated retention exception kept, each one listed, the evidence copy among them. Retail use of the pilot was paused, and it stays paused until question 2 of the readiness check is a yes. The response record assigns two decisions. The head of revenue chooses how the assistant will reach public rates: a connector to the revenue management system, or a dated reference copy in a library pilot users can open. The head of sales decides which corporate conditions in the CRM retail users may see. Retail use resumes only when the account manager's question, asked again, returns 236.00 euros and cites that source. The access and the link were removed, the Revenue site was given an owner, the group was scoped, and every reader but the incident team lost access to the workbook. That, not an archive tier, is what took it out of search: Microsoft documents that Copilot "only surfaces organizational data to which individual users have at least view permissions". Then came the test that matters: a search as the same user, after the index updated, that no longer returned the workbook. Last, the assistant's data protection impact assessment was updated. Article 35 requires an assessment before processing "likely to result in a high risk", and its paragraph 11 requires a review "at least when there is a change of the risk represented by processing operations". The group's assessment treated an assistant that had shown it could reach a retired file with personal data as such a change.

> *Do not fix the file first. Keep the evidence, decide on the record, then retire the file four ways and prove it with a search as the same user.*

ERA CARD

Reader. The assistant user, tested last: the plan is finished when the same person's search no longer returns the retired file.

Security. Custom groups are reviewed by name; anyone links expire; retirement means location, permissions, label and register.

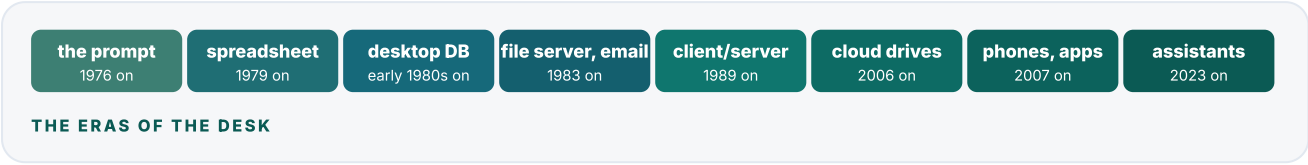
Cost. About 160 hours across eight people for one incident (illustrative), against about 276 hours, once, for the costed work of Chapter Fourteen, of which the top 20 clean-up is 120.

AI. The assistant's citation and stored prompt are evidence; keep them before you change anything, and retest as the user.

Monday rule: run a permissions report on your three largest custom groups, and for each one name an owner who must say what it should reach.

CLOSING

Epilogue: The Copy Outlives the Desk



THE ERAS OF THE DESK

In 1987 a reservations manager typed the season's rates into a Lotus 1-2-3 worksheet, saved it as RATES.WK1 on a floppy disk, printed it and faxed the printout to 12 hotels. The PC she used is gone. So are the floppy, the NetWare server that held RATES94.XLS, the Windows file server that replaced it and the on-premises CRM the CORP tab once queried. The desk itself has changed shape three times. The copies did not go anywhere. In March 2026 one of them, last saved in June 2022, answered a question for an account manager who had never heard of it, through an assistant that could read it and could not read the system that replaced it. Every desk in this book was retired. The copies outlived them all.

Six names, one workbook, no key

Counting its paths, the rate sheet had six names between 1987 and 2019: RATES.WK1 on a floppy disk, S:\REVENUE\RATES94.XLS, RATES98 final GM.xls, Rates master.xlsx on the file server, Revenue/Rates/Rates master.xlsx in a SharePoint Online library, and rates final GM v2.xlsx in a OneDrive. It never had a key. Its identity was a path, and a path is a locator. Each move was sensible on the day it was made: Excel replaced 1-2-3, email replaced the floppy, the cloud library replaced the file server. Each move also dropped something the next reader needed. The uplift macro stopped deriving rates in 1994. The festival blackout stayed in a typed memo in 1987 and was lost again in the 2022 import. A query that died in 2010 left the CORP tab looking like a live extract in 2026, its contacts as they were in 2010 and its rates as typed by hand until June 2022. Most people in the thread followed the practice of their day. They solved the problem in front of them, and nobody was asked to write down which copy was the answer.

That is the pattern the book has followed. The personal computer made the record a file a person could create, copy and rename, and made every copy editable, ownerless and indistinguishable from the original. Forty years of desk tools made copies cheaper, then added links, and never recorded a file's owner, meaning, authority or audience. The test the book set at the start still stands: for one important record, can you count its copies and name who can open each one?

The new reader is also a writer

The assistant in the Prologue was a reader. It did nothing it was not permitted to do. Microsoft documents that Copilot "only surfaces organizational data to which individual users have at least view permissions", and Google documents that Gemini "has the same access to Workspace data as you do". Both statements are the design working as described, and both are the reason Chapters Fourteen and Fifteen exist: an assistant inherits every permission its user has, including the ones nobody meant to give.

The same assistants now write. Microsoft documents that Copilot in Word drafts new content from a prompt and from a file the user picks, and that the user can keep, discard or regenerate the draft. Every kept draft is a new file, born as a working copy, with an author, a date and no authority state. If it quotes a rate, it is the twenty-fourth copy of the rate sheet. Microsoft also documents that it stores each prompt and response, including the citations used to ground it; that record was the first piece of evidence the group kept in March 2026.

OUR VIEW, AS OF NOVEMBER 2026

What changes. Three things. First, the forgotten copy is found. A workbook last saved in 2022 was forgotten by its owners and fully visible to an index; assistants read the long tail of a file estate that humans had stopped reading years ago. Second, copies multiply faster. A person who drafts with an assistant produces new files at the speed of a prompt, and each one starts life with no owner beyond its author and no state beyond "working copy". Third, the citation becomes evidence. An answer that names its source file starts a Ledger row for free: it turns "which copy did the answer use?" from an investigation into a lookup, though the copy's state, owner and reach still have to be recorded.

What does not change. Authority is not in the file. No index, label or model can tell from Rates master.xlsx alone that the revenue management system replaced it; only a recorded decision can. A copy still needs an owner, and the owner still leaves, moves or forgets. Deleting a record still means finding every place it went, and an assistant's chat history is one more place. Permission is still not purpose: being allowed to open a file never meant being meant to use it. And people, not tools, still decide which copy is the answer.

What I expect. Teams that keep an authority register and a Ledger for their top files will switch assistants on wider and sooner, because they can answer "where did that come from?" before a client asks. Teams that do not will find their retired files the way this group did, through a customer quoting an old price back to them. I do not expect assistants to fix stale or overshared files on their own. I expect them to make those files expensive sooner, which is a reason to do the work this book describes before the next pilot, not after it.

Back to the floppy disk

When the revenue manager walked the answer back, the trail ended on a floppy disk that no longer exists and a typed memo that nobody kept. The festival blackout, the one condition that would have stopped the 189.00 euro answer, was written in 1987 on paper, beside the sheet, never in it. In 2026 it lives as a clause in each corporate agreement in the CRM, which no assistant in the group could reach. The meaning outlived the desk too. It simply never travelled with the copy.

That is not an argument against spreadsheets. They are how the business adapts, and the rate sheet lasted 35 years because the revenue team could change pricing rules faster than any vendor could. It is an argument for writing down three things for every file that feeds a price, a regulator, a model or an assistant: which copy is the answer, who owns it, and who can open each of the others. The desk will change again. The copies will still be there.

> *Desks are replaced. Copies are inherited. Name the copy that answers, before the next reader asks.*

ERA CARD

Reader. From a fax machine in 1987 to an assistant in 2026; the next reader will also read what it is permitted to open.

Security. Inherited permission is the risk; groups, links and retired files are the controls.

Cost. Storage became cheap over 39 years; reconciliation, rework and one wrong answer did not.

AI. The assistant reads and now writes; give its drafts a state and its answers a Ledger row.

Monday rule: ask your assistant pilot one question whose answer you know comes from a replaced system, and see which copy it cites.

APPENDICES

APPENDIX A

The Ledger for Files: Template, Inventory and Readiness Check

The working tools of this book on a few pages: the blank Data Lineage Ledger with its eight loss codes, the five optional columns for files, the seven authority states, the desk file inventory worksheet from Chapter Fourteen, the assistant readiness check, and one worked row from the thread. The full Ledger for the rate sheet closes Chapter Thirteen, the end of Part V, and is not repeated here; it already includes the response (hop 10b) that Chapter Fifteen tells.

The blank Ledger

The Data Lineage Ledger: one row per hop

Record or dataset, hop	From	To	Mechanism and schedule	Format change; what was lost or assumed	Accountable owner	Who can change it	Classification; retention or erasure duty	Read by (people, apps, models, agents)	Cost per month; basis

Loss codes

MEANING	a code or definition did not travel with the value
CONTEXT	notes or surrounding knowledge stayed behind
PRECISION	type or rounding changed the value
TIME	century, time zone, or event versus processing time changed
IDENTITY	the key was renumbered, merged or split
OWNERSHIP	a copy has no accountable owner
RIGHTS	purpose, consent or legal basis did not travel
COMPLETENESS	rows or fields were dropped

Write one row per hop: each move or change after which a different reader holds the record; a deletion or a fix that changes who can read it is a hop too. Start at the point of use and walk back. A row with no loss is valid; write "no material loss". When an assistant writes a draft or a file, the write is a hop with its own row.

Optional columns for files

The ten core columns fit any record. Files need five more facts, because a file's identity is a path and its audience is a setting. Add them to column 1 (record) and column 9 (read by), or as extra columns.

Column	Write	Example (hop 7)
Path or URL	Where this copy opens today; a locator, not an identity	Revenue/Rates/Rates master.xlsx
Copies known	The count from the last census, and its date and limits	23 in March 2026, library versions not counted
Sharing scope	Person, group (with member count), organization, or anyone with the link	Group: All staff (internal), about 9,000
Authority state	One of the seven states below, for the decision this row serves	None recorded: retired only by an emailed memo
Last opened	Date from the platform's activity data, or "unknown" with the reason	Unknown: last saved June 2022; audit records reach back only months

The seven authority states

State	Meaning	Who may quote it
Authoritative	The source the owner names for one decision	Anyone, for that decision
Operational copy	A copy a system uses, refreshed from the source on a stated schedule	The system, within its refresh window
Reference copy	A read-only snapshot with an as-of date	Anyone, with the date
Working copy	A draft or analysis owned by its author	Nobody, as the answer
Retired	Replaced; marked by location, permissions, label and register	Nobody, for decisions
Archived	Kept for a duty; only its named readers can open or find it (an archive tier alone does not remove it from search)	Named readers with that duty
Disposed	Destroyed under a recorded decision; only the register entry remains	Nobody; the entry is the proof

The desk file inventory worksheet

One row per file found. Impact counts five tests (feeds a system of record, a regulator, a price, a model, an assistant); exposure counts four (shared beyond the team, personal data, more than one known copy, owner absent). Files at 2 or more on both go on the candidate list, and so does any file with personal data behind an anyone link; sort by exposure, then impact, then readers, and take the top 20.

File and path	Decision it serves; authoritative source	Impact (0 to 5)	Exposure (0 to 4)	Readers	Owner	State	Action and date
Rates master.xlsx, Revenue library	Public rate: revenue management system; corporate rate: CRM agreement	2	4	About 9,000	None since 2023	None recorded (memo only)	Owner, retire four ways, restrict; day 45

The assistant readiness check

Ten questions to answer before a workplace assistant is switched on for a group of users, or widened to a new one. Each question is tied to the hop where the rate sheet failed it. Questions 2 and 3 are gates: do not widen the pilot until both are yes. Every other "no" is a named risk with an owner and a date.

#	Question	Hop
1	For the decisions these users ask about most, is the authoritative source named in an authority record?	7
2	Can the assistant reach that source, through a connector or a dated export, under these users' permissions?	9
3	Are replaced files retired four ways (location, permissions, label and a register entry), not by memo?	7
4	Has every group that grants read on content in the index been reviewed by name and member count, custom and HR-fed groups included, not only the built-in everyone groups?	5, 5b, 9
5	Does every site and library in the index have a current owner?	7
6	Are anyone links and personal-drive copies of governed files listed, and expired or removed on sites that hold personal data?	6
7	Do files in the index hold personal data their readers have no purpose to see, such as a pasted contact or guest tab?	4
8	Do tabs that look like live extracts still refresh from their source, and do they show their owner and last refresh?	4b

#	Question	Hop
9	Can an erasure be traced to every copy the assistant can read: versions, drives, mail, histories?	6b
10	Is there a rule for where answers and work data may go (client email, personal chat accounts), and are prompts, answers and citations kept long enough to trace an answer to its copy and state?	8, 10

One worked row

Hop 7, the retirement that was not, as the Ledger records it. Read the columns against the optional file columns above: the owner is "nobody for the workbook after 2023", the readers are everyone in All staff (internal), and the state lives in an email. The row is the evidence Chapter Fifteen used to retire the file properly.

THE HOP · 7 · 2022 · RATES MASTER.XLSX TO RENTED REVENUE MANAGEMENT SYSTEM (ONE IMPORT)

Record	The rate sheet (hotel and retail group), hop 7, 2022
From	Rates master.xlsx
To	Rented revenue management system (one import)
Mechanism and schedule	Public rates are imported in March 2022 and then maintained in the new system (236.00 euros: the harbor hotel's public rate for the 2026 festival dates); the workbook is kept in parallel until its last save in June 2022; in July 2022 a memo emailed to the revenue team retires it
Format change; what was lost or assumed	COMPLETENESS OWNERSHIP System import of public rates. **The import carries public rates, not the CORP tab or its conditions (the festival blackout is lost a second time); the workbook is not archived, relabeled or removed; the memo is an email, so no file anyone can search says the workbook is retired; separately, in 2023 the site owner leaves and nobody is assigned**
Accountable owner	Head of revenue (the system); nobody for the workbook after 2023
Who can change it	The revenue team (edit) and site collection administrators; nobody is responsible for it
Classification; retention or erasure duty	Personal data and commercial terms, retired by memo only
Read by (people, apps, models, agents)	Anyone in All staff (internal)
Cost per month; basis	Per month: the system subscription; the workbook costs nothing until someone reads it (illustrative)

APPENDIX B

Timeline

Twenty-five dated points from the book, each checked against a primary source or an institutional history. Where only a month or a year is certain, only that is given. The thread's own dates are invented and are not listed. Full references are in Appendix E.

Date	Event	Ch	Source
1976	Digital Research releases CP/M	1	*The MS-DOS Encyclopedia*, Section I
1979	Bricklin and Frankston develop VisiCalc	3	Computer History Museum timeline
12 Aug 1981	IBM announces the IBM Personal Computer; its operating system, MS-DOS, comes from Microsoft (IBM calls it IBM Personal Computer DOS)	1	IBM History, "The IBM PC"; *The MS-DOS Encyclopedia*, Section I
About Sep 1985	Lotus 1-2-3 Release 2 ships; it writes the WK1 format	3	CHM, Lotus 1-2-3 workshop transcript
30 Sep 1985	Excel for the Macintosh ships to retail stores	3	Microsoft, History of Microsoft 1985
19 Nov 1987	Excel for Windows ships	3	Microsoft, History of Microsoft 1987
3 May 1989	Microsoft and Ashton-Tate announce shipment of SQL Server 1.0	8	Microsoft, History of Microsoft 1989
1992	Microsoft releases the Open Database Connectivity (ODBC) Software Development Kit 1.0	8	Microsoft, History of Microsoft 1992
16 Nov 1992	Microsoft announces availability of Access	5	Microsoft, History of Microsoft 1992
24 Oct 1995	Directive 95/46/EC on the protection of personal data adopted	8	EUR-Lex, CELEX 31995L0046
28 Mar 2001	Microsoft announces SharePoint Portal Server ready for release	9	Microsoft News, 28 Mar 2001
Oct 2005	RFC 4180 describes a common CSV format (Informational)	7	RFC Editor, RFC 4180
Dec 2006	Ecma approves ECMA-376, Office Open XML, first edition	2	Ecma International, ECMA-376

Date	Event	Ch	Source
28 Jun 2011	Office 365 launches, including SharePoint Online	9	Microsoft News, 28 Jun 2011
24 Apr 2012	Google introduces Google Drive	9	Official Google Blog
16 Jan 2013	JPMorgan Chase publishes its Management Task Force report on the 2012 CIO losses	4	JPMorgan Chase Form 8-K
19 Feb 2014	OneDrive, the renamed SkyDrive, launches worldwide	9	Microsoft News, 19 Feb 2014
23 Aug 2016	Ziemann, Eren and El-Osta find gene name errors in 19.6 percent of papers with Excel gene lists	4	*Genome Biology* 17:177
25 May 2018	The GDPR applies (Article 99)	11	Regulation (EU) 2016/679
4 Oct 2020	Public Health England reports 15,841 cases left out of daily figures	4	GOV.UK, PHE statement
8 Aug 2023	PSNI spreadsheet with a hidden worksheet published in an FOI response	12	ICO penalty notice, 26 Sep 2024
1 Nov 2023	Microsoft 365 Copilot generally available to enterprise customers	13	Microsoft Tech Community, 1 Nov 2023
Sep 2025	NIST SP 800-88 Rev. 2 replaces Rev. 1 on media sanitization	11	NIST CSRC
17 Apr 2026	SR 26-2 supersedes SR 11-7 on model risk management	4	Federal Reserve, SR 26-2
31 Jul 2026	New enablement of Restricted SharePoint Search blocked	13, 15	Microsoft Learn

APPENDIX C

Failure Library

The sourced failure cases in this book, each with the report, paper or penalty notice behind it. Only facts found in those documents are given. The mechanism column is the author's reading through the Ledger's loss codes, not a finding of the report. Each case is told once, in the chapter named; this table points to it.

Case, year	Mechanism (loss code)	Report and where to find it	Ch
JPMorgan Chase CIO losses, 2012	PRECISION OWNERSHIP The new value-at-risk model "operated through a series of Excel spreadsheets", completed manually "by a process of copying and pasting data from one spreadsheet to another" (p. 124); one spreadsheet "divided by their sum instead of their average" (p. 128). A model of record ran on hand-fed desk files.	*Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses*, 16 January 2013, announced in Form 8-K the same day. sec.gov/Archives/edgar/data/19617/000001961713000077/ ; elischolar.library.yale.edu/ypfs-documents/454	4
Reinhart and Rogoff replication, 2013	COMPLETENESS PRECISION Herndon, Ash and Pollin replicated a published result on public debt and growth and found "coding errors, selective exclusion of available data, and unconventional weighting of summary statistics". A published figure rested on a spreadsheet nobody outside the authors had checked.	Thomas Herndon, Michael Ash and Robert Pollin, "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff", PERI Working Paper 322, University of Massachusetts Amherst, April 2013. peri.umass.edu	4
Gene symbols converted to dates, 2016	MEANING Screening supplementary Excel files from 18 journals published 2005 to 2015, the authors confirmed gene name errors in 987 files from 704 papers: 19.6 percent of papers with Excel gene lists. A symbol was read as a date when a file was opened.	Mark Ziemann, Yotam Eren and Assam El-Osta, "Gene name errors are widespread in the scientific literature", *Genome Biology* 17:177, 23 August 2016. doi.org/10.1186/s13059-016-1044-7	4
Public Health England case data, 2020	COMPLETENESS "15,841 cases between 25 September and 2 October were not included in the reported daily COVID-19 cases." PHE said the cause was that "some files containing positive test results exceeded the maximum file size". The statement names no product and no row limit.	"PHE statement on delayed reporting of COVID-19 cases", GOV.UK, 4 October 2020, updated 5 October 2020. gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases	4

Case, year	Mechanism (loss code)	Report and where to find it	Ch
PSNI hidden worksheet, 2023	RIGHTS On 8 August 2023 a spreadsheet released in response to a freedom of information request disclosed the personal data of 9,483 police officers and staff (para 6 to 7); the data sat on a second worksheet, "SAP DOWNLOAD", that the people checking the file did not see, and it was public for about 2 hours and 20 minutes (para 21). The FOI instruction had no guidance on checking Excel files for hidden data (para 38), and training did not raise the risk (para 44). Penalty: 750,000 pounds, reduced from 5.6 million under the public sector approach.	Information Commissioner's Office, *Penalty Notice: Police Service of Northern Ireland*, 26 September 2024, published 3 October 2024. ico.org.uk/action-weve-taken/enforcement/2024/10/police-service-of-northern-ireland-mpn/ ; it cites the PSNI independent review final report of 11 December 2023	12

How to use the library

Read the cases by mechanism, not by industry. The JPMorgan and replication cases are desk files acting as models of record with no second pair of eyes; they score high on Chapter Fourteen's impact test, because each fed a decision far beyond its author's desk, and even below the exposure line such a file gets an owner and an authority record, as the close workbook in that chapter's ranking does. The gene symbol and PHE cases are what a file format or a tool does to a value when nobody states the type or the limit; Chapter Seven's Type Contract is built for them. The PSNI case is the mirror of the rate sheet's CORP and VIP tabs: personal data in part of a workbook the sender did not look at, reaching readers it was never meant for. Then ask question 7 of the readiness check in Appendix A, the one the PSNI case would have failed, of your own files.

APPENDIX D

Glossary

Terms used in this book, in plain words, as of November 2026.

Term	Meaning
8.3 name	An MS-DOS file name: up to eight characters, a period, and up to three for the extension.
Access control list (ACL)	The list of who may do what to a file, kept by NTFS for each file and folder.
Anyone link	A sharing link that opens a file without signing in; whoever holds it can use it.
Archived	An authority state: kept for a duty, with access limited to the readers that duty names, so it is out of everyone else's search and everyday access. Moving a file to an archive tier does not by itself remove it from search.
Authoritative	An authority state: the one source the owner names for a decision.
Authority record	For one decision: the authoritative source, its owner, the date and the state of every other copy.
Authority register	The place authority records and state changes are kept, readable by people and tools.
Authority state	One of seven states a copy holds for a decision, from authoritative to disposed.
BIFF	Binary Interchange File Format, the binary layout of Excel's .xls workbooks.
Copy census	A count of every place a record can be opened today, with sharing scope and state.
CSV	Comma-separated values: plain text rows; RFC 4180 (2005) describes a common form, not a binding standard.
Custom group	A group an organization creates, as opposed to a built-in one such as "Everyone except external users".
Data Lineage Ledger	A table with one row per hop, ten core columns and any loss codes.
Data protection impact assessment (DPIA)	The assessment GDPR Article 35 requires before processing likely to result in a high risk.
Desk file inventory	A list of files scored for impact (0 to 5) and exposure (0 to 4) to pick the top 20.
Disposed	An authority state: destroyed under a recorded decision; only the register entry remains.
Erasure fan-out	Tracing an erasure duty from the record to every copy it reached, with proof for each.
Exception	A file kept outside the rules, with an owner, the criteria and a review date.
Exposure	In the inventory: sharing beyond the team, personal data, more than one copy, owner absent.
Hop	A move, change, deletion or fix after which different readers hold the record.

Term	Meaning
Impact	In the inventory: the file feeds a system of record, a regulator, a price, a model or an assistant.
Index	The search service's store of content and permissions that an assistant searches.
Leaver, mover rules	What happens to a person's files and sites when they leave or change role.
Legal hold	A duty to keep records that may be evidence, suspending normal deletion.
Linked table	A table in a desktop database that reads another file by its path; it breaks when the path moves.
Loss code	One of eight Ledger labels: MEANING, CONTEXT, PRECISION, TIME, IDENTITY, OWNERSHIP, RIGHTS, COMPLETENESS.
Macro	Code stored in a workbook or document that runs inside the desk tool.
ODBC	Open Database Connectivity, a standard interface a desk tool uses to query a database.
Office Open XML	The zipped XML format of .xlsx and .docx files (ECMA-376, ISO/IEC 29500).
Operational copy	An authority state: a copy a system uses, refreshed from the source on a schedule.
Owner (accountable)	A person who approves changes to a file and answers when it is wrong.
Path	Where a file opens: a locator, not an identity.
Personal data breach	In the GDPR (Article 4(12)), a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data.
Readiness check	Ten questions to answer before an assistant is switched on for a group of users.
Reference copy	An authority state: a read-only snapshot with an as-of date.
Retired	An authority state: replaced, not for decisions; marked by location, permissions, label and register.
Sharing scope	Who a copy is open to: a person, a group, the organization, or anyone with the link.
Supervisory authority	The public body that enforces data protection law, notified of breaches under GDPR Article 33.
Trustee rights	NetWare's permissions on a directory or file for a user or group.
Type Contract	This series' seven-line box stating a value's type, encoding, unit, time basis, identity scope, absence and authority.
Version history	Earlier versions a library keeps for an item; not a backup, and in reach of an erasure duty.
Working copy	An authority state: a draft or analysis owned by its author; never quoted as the answer.
Workplace assistant	An AI assistant that answers from an organization's files and mail within each user's permissions.

APPENDIX E

Sources

Every dated fact in this book was checked against the source listed for its chapter on 19 and 20 November 2026. Primary sources come first: original papers, standards, statutes, judgments, official reports, vendor announcements and institutional histories. This list holds 243 sources; the claim-by-claim record behind it is kept with the book and is available on request.

Reader's Brief: One Sheet, Six Questions

- Illustrative running example of this book; invented

One Sheet, Since 1987

- Illustrative running example of this book; invented
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy

Chapter 1. The Command Line and the File

- IBM History, "The IBM PC" https://www.ibm.com/history/personal-computer
- The MS-DOS Encyclopedia, Section I (Microsoft Press, 1988) https://www.pcjs.org/documents/books/mspl13/msdos/encyclopedia/section1/
- The MS-DOS Encyclopedia, Section II https://www.pcjs.org/documents/books/mspl13/msdos/encyclopedia/section2/
- Microsoft Learn, [MS-FSCC] 2.1.5.2.1 "8.3 Filename" https://learn.microsoft.com/en-us/openspecs/windows_protocols/ms-fscc/18e63b13-ba43-4f5f-a5b7-11e871b71f14
- Microsoft Learn, "Naming Files, Paths, and Namespaces" https://learn.microsoft.com/en-us/windows/win32/fileio/naming-a-file
- Microsoft Learn, "Overview of FAT, HPFS, and NTFS File Systems" (KB 100108) https://learn.microsoft.com/en-us/troubleshoot/windows-client/backup-and-storage/fat-hpfs-and-ntfs-file-systems
- Microsoft Learn, Windows commands, "copy" https://learn.microsoft.com/en-us/windows-server/administration/windows-commands/copy
- Microsoft Learn, Windows commands, "ren" https://learn.microsoft.com/en-us/windows-server/administration/windows-commands/ren
- Computer History Museum, "PC Software Workshop: Lotus 1-2-3" (recorded 6 May 2004, CHM X4310.2008) https://archive.computerhistory.org/resources/access/text/2012/04/102658156-05-01-acc.pdf
- Computer History Museum, Timeline of Computer History: Software and Languages https://www.computerhistory.org/timeline/software-languages/
- Novell press room, "Novell History" (Japanese edition) https://www.novell.com/ja-jp/news/press/pressroom/history.html
- Microsoft Learn, "The History of Microsoft - 1985" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1985
- Microsoft Learn, "The History of Microsoft - 1987" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1987
- Microsoft Learn, "The History of Microsoft - 1989" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1989
- Microsoft Learn, "The History of Microsoft - 1992" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1992
- RFC Editor, RFC 4180 https://www.rfc-editor.org/info/rfc4180
- Ecma International, ECMA-376 https://ecma-international.org/publications-and-standards/standards/ecma-376/

- Official Google Blog, "Introducing Google Drive… yes, really" https://googleblog.blogspot.com/2012/04/introducing-google-drive-yes-really.html
- Microsoft Tech Community, "Microsoft 365 Copilot is generally available" https://techcommunity.microsoft.com/blog/microsoft-copilot-blog/microsoft-365-copilot-is-generally-available/3969331
- Illustrative running example of this book; invented
- Microsoft Knowledge Base 299648, Description of NTFS date and time stamps for files and folders (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/299/648.HTM
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy

Chapter 2. What a File Knows and What It Forgets

- The MS-DOS Encyclopedia, Section II (Microsoft Press, 1988) https://www.pcjs.org/documents/books/mspl13/msdos/encyclopedia/section2/
- Microsoft Learn, "File System Functionality Comparison" https://learn.microsoft.com/en-us/windows/win32/fileio/filesystem-functionality-comparison
- Novell press room, "Novell History" (Japanese edition) https://www.novell.com/ja-jp/news/press/pressroom/history.html
- Novell, ConsoleOne 1.3.x User Guide, 4.5 "About NetWare Rights" https://www.novell.com/documentation/consol13/c1_enu/data/hrv09hes.html
- Microsoft Learn, "Security Descriptors" https://learn.microsoft.com/en-us/windows/win32/secauthz/security-descriptors
- Microsoft Support, "View or change the properties for an Office file" https://support.microsoft.com/en-us/office/view-or-change-the-properties-for-an-office-file-21d604c2-481e-4379-8e54-1dd4622c6b75
- Microsoft Learn, .NET API, "PackageProperties Class (System.IO.Packaging)" https://learn.microsoft.com/en-us/dotnet/api/system.io.packaging.packageproperties
- Microsoft Learn, Windows commands, "copy" https://learn.microsoft.com/en-us/windows-server/administration/windows-commands/copy
- Microsoft Graph v1.0, "driveItem resource type" https://learn.microsoft.com/en-us/graph/api/resources/driveitem?view=graph-rest-1.0
- Computer History Museum, "PC Software Workshop: Lotus 1-2-3" (recorded 6 May 2004) https://archive.computerhistory.org/resources/access/text/2012/04/102658156-05-01-acc.pdf
- Microsoft Learn, [MS-XLS] Glossary https://learn.microsoft.com/en-us/openspecs/office_file_formats/ms-xls/24fed501-1f2e-435b-a371-ad29a57dba58
- Ecma International, ECMA-376 https://ecma-international.org/publications-and-standards/standards/ecma-376/
- ISO news, "ISO/IEC DIS 29500 receives necessary votes for approval as an International Standard" (April 2008) https://www.iso.org/news/2008/04/Ref1123.html
- ISO, "ISO/IEC 29500-1:2008" https://www.iso.org/standard/51463.html
- ISO, "ISO/IEC 26300:2006" https://www.iso.org/standard/43485.html
- Microsoft Learn, "Naming Files, Paths, and Namespaces" https://learn.microsoft.com/en-us/windows/win32/fileio/naming-a-file
- The MS-DOS Encyclopedia, Section I https://www.pcjs.org/documents/books/mspl13/msdos/encyclopedia/section1/
- Microsoft Support, "Enable and configure unique Document IDs" https://support.microsoft.com/en-us/office/enable-and-configure-unique-document-ids-ea7fee86-bd6f-4cc8-9365-8086e794c984
- Microsoft Learn, "Overview of FAT, HPFS, and NTFS File Systems" (KB 100108) https://learn.microsoft.com/en-us/troubleshoot/windows-client/backup-and-storage/fat-hpfs-and-ntfs-file-systems
- Illustrative running example of this book; invented
- Microsoft Support, File formats that are supported in Excel https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247

Chapter 3. VisiCalc, Lotus 1-2-3, Excel: the Grid as a System

- Computer History Museum, Timeline of Computer History: Software and Languages, 1979 https://www.computerhistory.org/timeline/software-languages/
- Computer History Museum, Revolution, Software: Putting PCs to Work https://www.computerhistory.org/revolution/personal-computers/17/305
- Computer History Museum, Mitchell Kapor profile https://computerhistory.org/profile/mitchell-kapor/

- Computer History Museum, PC Software Workshop: Lotus 1-2-3 (CHM X4310.2008) https://archive.computerhistory.org/resources/access/text/2012/04/102658156-05-01-acc.pdf
- Microsoft Learn, The History of Microsoft, 1985 https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1985
- Microsoft Learn, The History of Microsoft, 1987 https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1987
- Microsoft Learn, Excel incorrectly assumes that the year 1900 is a leap year https://learn.microsoft.com/en-us/troubleshoot/microsoft-365-apps/excel/wrongly-assumes-1900-is-leap-year
- Microsoft Knowledge Base 269178, XL97: Macro Compatibility with Microsoft Excel (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/269/178.HTM
- Microsoft Knowledge Base 119141, "Macro Error..." with Lotus File Combine Macro (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/119/141.HTM
- Microsoft Knowledge Base Q120733, Error Running Lotus 1-2-3 Macros with 3-D References (archived copy) https://techshelps.github.io/MSDN/KB/Source/excel/q120733.htm
- Microsoft Knowledge Base Q111842, XL5 Err Msg: "That Name Is not Valid" with Macro Translator (archived copy) https://techshelps.github.io/MSDN/KB/Source/excel/q111842.htm
- Microsoft Knowledge Base 258777, XL2000: Some Macros Don't Run After You Upgrade to Excel 2000 (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/258/777.HTM
- Microsoft Learn, [MS-XLS] Glossary https://learn.microsoft.com/en-us/openspecs/office_file_formats/ms-xls/24fed501-1f2e-435b-a371-ad29a57dba58
- Microsoft Support, More on data validation https://support.microsoft.com/en-us/office/more-on-data-validation-f38dee73-9900-4ca6-9301-8a5f6e1f0c4c
- Microsoft Support, Protect a worksheet https://support.microsoft.com/en-us/office/protect-a-worksheet-3179efdb-1285-4d49-a9c3-f4ca36276de6
- Microsoft Support, View previous versions of Office files https://support.microsoft.com/en-us/office/view-previous-versions-of-office-files-5c1e076f-a9c9-41b8-8ace-f77b9642e2c2
- Microsoft Support, Use Excel with earlier versions of Excel https://support.microsoft.com/en-us/excel/use-excel-with-earlier-versions-of-excel
- Microsoft Support, Excel specifications and limits https://support.microsoft.com/en-us/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3
- Illustrative running example of this book; invented
- Microsoft Support, "Excel file is locked for editing" https://support.microsoft.com/office/excel-file-is-locked-for-editing-6fa93887-2c2c-45f0-abcc-31b04aed68b3
- Microsoft Support, "Date systems in Excel" https://support.microsoft.com/office/date-systems-in-excel-e7fe7167-48a9-4b96-bb53-5612a800b487
- Microsoft Support, File formats that are supported in Excel https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247
- European Central Bank, Introduction of the euro https://www.ecb.europa.eu/euro/intro/html/index.en.html

Chapter 4. When the Spreadsheet Is the System

- Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses (16 January 2013), copy in the Yale Program on Financial Stability repository https://elischolar.library.yale.edu/ypfs-documents/454
- Same report
- Herndon, Ash and Pollin, Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff, PERI Working Paper 322 https://peri.umass.edu/images/WP322.pdf; publication page https://peri.umass.edu/publication/does-high-public-debt-consistently-stifle-economic-growth-a-critique-of-reinhart-and-rogoff/
- GOV.UK, PHE statement on delayed reporting of COVID-19 cases https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases
- Ziemann M, Eren Y, El-Osta A, Gene name errors are widespread in the scientific literature https://doi.org/10.1186/s13059-016-1044-7
- HGNC, Summer newsletter 2020 https://blog.genenames.org/newsletters/2020/08/28/Summer_newsletter/
- https://doi.org/10.1038/s41588-020-0669-3
- Microsoft Support, Set automatic data conversions https://support.microsoft.com/en-us/excel/set-automatic-data-conversions

- Computer History Museum, PC Software Workshop: Lotus 1-2-3 (6 May 2004, CHM X4310.2008) https://archive.computerhistory.org/resources/access/text/2012/04/102658156-05-01-acc.pdf
- Microsoft Learn (archived), Protecting Office Documents from Macro Viruses (Office 2000) https://learn.microsoft.com/en-us/previous-versions/office/developer/office2000/aa189865(v=office.10)
- Microsoft Learn, Macros from the internet are blocked by default in Office https://learn.microsoft.com/en-us/microsoft-365-apps/security/internet-macros-blocked
- Microsoft Support, Remove hidden data and personal information by inspecting documents, presentations, or workbooks https://support.microsoft.com/en-us/office/remove-hidden-data-and-personal-information-by-inspecting-documents-presentations-or-workbooks-356b7b5d-77af-44fe-a07f-9aa4d085966f
- Actuarial Standards Board, ASOP No. 56, Modeling https://www.actuarialstandardsboard.org/asops/modeling-3/
- Federal Reserve, SR 11-7 Attachment, Supervisory Guidance on Model Risk Management https://www.federalreserve.gov/boarddocs/srletters/2011/sr1107a1.pdf
- Federal Reserve, SR 26-2, Revised Guidance on Model Risk Management https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm
- Illustrative running example of this book; invented

Chapter 5. dBase and Access: the Desktop Database

- Computer History Museum, Information Technology Corporate Histories Project, Ashton-Tate https://archive.computerhistory.org/resources/access/text/2012/10/102746512-05-01-acc.pdf
- Microsoft Learn, The History of Microsoft, 1992 https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1992
- Microsoft Support, Add or change a table's primary key in Access https://support.microsoft.com/en-us/office/add-or-change-a-table-s-primary-key-in-access-07b4a84b-0063-4d56-8b00-65f2975e4379
- Microsoft Support, Ways to share an Access desktop database https://support.microsoft.com/en-US/Access/ways-to-share-an-access-desktop-database
- Microsoft Support, Split an Access database https://support.microsoft.com/en-us/office/split-an-access-database-3015ad18-a3a1-4e9c-a7f3-51b1d73498cc
- Microsoft Support, Access specifications https://support.microsoft.com/en-us/office/access-specifications-0cf3c66f-9cf2-4e32-9568-98c1025bb47c
- Microsoft Support, Manage linked tables https://support.microsoft.com/en-us/office/manage-linked-tables-1d9346d6-953d-4f85-a9ce-4caec2262797
- Microsoft Support, Import or link to data in an Excel workbook https://support.microsoft.com/en-us/office/import-or-link-to-data-in-an-excel-workbook-a1952878-7c58-47b1-893d-e084913cc958
- Microsoft Knowledge Base 149051, ACC95: "Import or Link Spreadsheet" Help Topic Error (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/149/051.HTM
- Illustrative running example of this book; invented
- Microsoft Knowledge Base 904018, Description of the update for Access 2002: October 18, 2005 (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/904/018.HTM
- Microsoft Knowledge Base 157495, Odd Workbook Behavior After Modification by MS Access (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/157/495.HTM
- Microsoft Support, Discontinued features and modified functionality in Access 2013 https://support.microsoft.com/en-us/office/discontinued-features-and-modified-functionality-in-access-2013-bc006fc3-5b48-499e-8c7d-9a2dfef68e2f
- Microsoft Knowledge Base 904953, You cannot change, add, or delete data in tables that are linked to an Excel workbook in Office Access 2003 or in Access 2002 (archived) https://ftp.zx.net.nz/pub/archive/ftp.microsoft.com/MISC/KB/en-us/904/953.HTM

Chapter 6. Shared Drives, Email and the Copy as Default

- Novell press room, Novell History (page still served, Micro Focus copyright) https://www.novell.com/ja-jp/news/press/pressroom/history.html
- Microsoft Tech Community, 30 years of Windows Server https://techcommunity.microsoft.com/blog/sbs/30-years-of-windows-server/3884810
- Microsoft Learn, What is SMB File Sharing for Windows and Windows Server https://learn.microsoft.com/en-us/windows-server/storage/file-server/file-server-smb-overview
- RFC Editor, RFC 2045 https://www.rfc-editor.org/info/rfc2045

- RFC Editor, RFC 1341 https://www.rfc-editor.org/info/rfc1341
- Microsoft Learn (archive), Office 97 Resource Kit, Chapter 12: Upgrading from Previous Versions of MS Excel https://learn.microsoft.com/en-us/previous-versions/tn-archive/cc749827(v=technet.10)
- Microsoft Support, File formats that are supported in Excel https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247
- Microsoft Learn, Naming Files, Paths, and Namespaces https://learn.microsoft.com/en-us/windows/win32/fileio/naming-a-file
- Illustrative running example of this book; invented

Chapter 7. CSV, the Format That Never Died

- Shafranovich, RFC 4180 (published as an RFC, Informational), Common Format and MIME Type for Comma-Separated Values (CSV) Files https://www.rfc-editor.org/rfc/rfc4180
- W3C, Model for Tabular Data and Metadata on the Web https://www.w3.org/TR/tabular-data-model/
- Microsoft Support, File formats that are supported in Excel https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247
- Microsoft Support, Import or export text (.txt or .csv) files https://support.microsoft.com/en-us/office/import-or-export-text-txt-or-csv-files-5250ac4c-663c-47ce-937b-339e391393ba
- Microsoft Learn, DateTimeFormatInfo.ShortDatePattern Property https://learn.microsoft.com/dotnet/api/system.globalization.datetimeformatinfo.shortdatepattern
- Author's computation, 20 November 2026
- Microsoft Support, Keeping leading zeros and large numbers https://support.microsoft.com/en-us/office/keep-leading-zeros-and-large-numbers-1bf7b935-36e1-4985-842f-5dfa51f85fe7
- Microsoft Support, Opening CSV UTF-8 files correctly in Excel https://support.microsoft.com/en-us/office/opening-csv-utf-8-files-correctly-in-excel-8a935af5-3416-4edd-ba7e-3dfd2bc4a032
- OWASP Foundation, CSV Injection https://community.owasp.org/attacks/CSV_Injection
- Illustrative running example of this book; invented

Chapter 8. Client/Server and the Pipe to the Desk

- Microsoft Learn, The History of Microsoft, 1989 https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1989
- Microsoft Learn, The History of Microsoft, 1992 https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1992
- Microsoft Learn, What Is ODBC? https://learn.microsoft.com/en-us/sql/odbc/reference/what-is-odbc
- Microsoft Support, Connection properties https://support.microsoft.com/en-us/office/connection-properties-9d3599a9-e9b3-461d-99b2-c5505ddae6e0
- EUR-Lex, Directive 95/46/EC https://eur-lex.europa.eu/eli/dir/1995/46/oj; legislation.gov.uk record https://www.legislation.gov.uk/eudr/1995/46/contents
- Microsoft Support, File formats that are supported in Excel https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247
- Microsoft Learn, What Is Power Query? https://learn.microsoft.com/en-us/power-query/power-query-what-is-power-query
- Microsoft Learn, Configure scheduled refresh (Power BI) https://learn.microsoft.com/en-us/power-bi/connect-data/refresh-scheduled-refresh
- Illustrative running example of this book; invented

Chapter 9. From Shared Drive to Cloud Drive

- Microsoft News, "Leading Systems Integrators Ready for the Availability of Microsoft SharePoint Portal Server" (28 March 2001) https://news.microsoft.com/source/2001/03/28/leading-systems-integrators-ready-for-the-availability-of-microsoft-sharepoint-portal-server/
- Microsoft News, "Microsoft Launches Office 365 Globally" (28 June 2011) https://news.microsoft.com/source/2011/06/28/microsoft-launches-office-365-globally/

- The Official Microsoft Blog, "Microsoft announces SkyDrive will soon be renamed OneDrive" (27 January 2014) https://blogs.microsoft.com/blog/2014/01/27/microsoft-announces-skydrive-will-soon-be-renamed-onedrive/
- Microsoft News, "Microsoft Office showcases future of work" (3 March 2014) https://news.microsoft.com/source/2014/03/03/microsoft-office-showcases-future-of-work-2/
- Official Google Blog, "Introducing Google Drive… yes, really" (24 April 2012) https://googleblog.blogspot.com/2012/04/introducing-google-drive-yes-really.html
- Microsoft Support, "How versioning works in lists and libraries" https://support.microsoft.com/en-us/office/how-versioning-works-in-lists-and-libraries-0f6cd105-974f-44a4-aadb-43ac5bdfd247
- Microsoft Learn, "Plan version storage for document libraries" (ms.date 2025-05-29) https://learn.microsoft.com/en-us/sharepoint/plan-version-storage
- Microsoft Support, "File storage in Microsoft Teams" https://support.microsoft.com/en-us/teams/files/file-storage-in-microsoft-teams
- Microsoft Learn, "Teams and SharePoint integration" (ms.date 2026-06-25) https://learn.microsoft.com/en-us/sharepoint/teams-connected-sites
- Microsoft Learn, "Manage sharing settings for SharePoint and OneDrive in Microsoft 365" (ms.date 2026-06-30) https://learn.microsoft.com/en-us/sharepoint/turn-external-sharing-on-or-off
- Microsoft Learn, "How shareable links work in OneDrive and SharePoint in Microsoft 365" (ms.date 2026-11-06) https://learn.microsoft.com/en-us/sharepoint/shareable-links-anyone-specific-people-organization
- Microsoft Learn, "Data access governance reports for SharePoint sites" (ms.date 2026-07-09) https://learn.microsoft.com/en-us/sharepoint/data-access-governance-reports
- Microsoft Support, "Collaborate on Excel workbooks at the same time with co-authoring" https://support.microsoft.com/en-us/excel/get-started/collaborate-on-excel-workbooks-at-the-same-time-with-co-authoring
- Microsoft Learn, "OneDrive retention and deletion" (ms.date 2026-06-02) https://learn.microsoft.com/en-us/sharepoint/retention-and-deletion
- Microsoft Learn, "Set the OneDrive retention for deleted users" (ms.date 2025-04-05) https://learn.microsoft.com/en-us/sharepoint/set-retention
- Microsoft Learn, "Manage ownerless Microsoft 365 groups and teams" (ms.date 2026-08-13) https://learn.microsoft.com/en-us/microsoft-365/admin/create-groups/ownerless-groups-teams
- Microsoft Learn, "Create a SharePoint site ownership policy" (ms.date 2026-05-14) https://learn.microsoft.com/en-us/sharepoint/site-ownership-policy
- Illustrative running example of this book; invented

Chapter 10. The Copy That Left the Building

- Microsoft Support, "Save disk space with OneDrive Files On-Demand for Windows" https://support.microsoft.com/en-us/office/save-disk-space-with-onedrive-files-on-demand-for-windows-0e6860d3-d9f3-4971-b321-7092438fb38e
- Microsoft Learn, "BitLocker device encryption" (ms.date 2026-11-03) https://learn.microsoft.com/en-us/windows/security/operating-system-security/data-protection/bitlocker/
- Apple Platform Security, "Data Protection overview" (published 3 August 2026) https://support.apple.com/guide/security/data-protection-overview-secf6276da8a/web
- Apple Platform Security, "Passcodes and passwords" (published 19 December 2024) https://support.apple.com/guide/security/passcodes-and-passwords-sec20230a10d/web
- Android Open Source Project, "Encryption" https://source.android.com/docs/security/features/encryption
- sqlite.org, "About SQLite" https://www.sqlite.org/about.html
- sqlite.org, "Most Widely Deployed and Used Database Engine" https://www.sqlite.org/mostdeployed.html
- sqlite.org, "History Of SQLite Releases" https://www.sqlite.org/chronology.html
- sqlite.org, "Temporary Files Used By SQLite" https://www.sqlite.org/tempfiles.html
- sqlite.org, "Pragma statements supported by SQLite" (secure_delete) https://www.sqlite.org/pragma.html
- Microsoft Learn, "Resolve sync issues in OneDrive for work or school" (ms.date 2025-01-03) https://learn.microsoft.com/en-us/troubleshoot/sharepoint/sync/troubleshoot-sync-issues
- Microsoft Support, "Collaborate on Excel workbooks at the same time with co-authoring" https://support.microsoft.com/en-us/excel/get-started/collaborate-on-excel-workbooks-at-the-same-time-with-co-authoring
- Microsoft Learn, "App protection policies overview" (ms.date 2026-03-04) https://learn.microsoft.com/en-us/intune/app-management/protection/overview
- eCFR, 45 CFR Part 164 Subpart D https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-D

- HHS, "Guidance to Render Unsecured Protected Health Information Unusable, Unreadable, or Indecipherable to Unauthorized Individuals" https://www.hhs.gov/hipaa/for-professionals/breach-notification/guidance/index.html
- HHS, "Breach Notification Rule" https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html
- Illustrative running example of this book; invented

Chapter 11. Deleting a File for Real

- Microsoft Learn, "Learn about retention for SharePoint and OneDrive" (ms.date 22 September 2025) https://learn.microsoft.com/en-us/purview/retention-policies-sharepoint
- Microsoft Learn, "Learn about retention policies and retention labels" (ms.date 22 July 2026) https://learn.microsoft.com/en-us/purview/retention
- Microsoft Learn, "Recoverable Items folder in Exchange Online" (ms.date 25 September 2026) https://learn.microsoft.com/en-us/exchange/security-and-compliance/recoverable-items-folder/recoverable-items-folder
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" (ms.date 9 July 2026) https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
- Microsoft Learn, "Learn about retention for Copilot and AI apps" (ms.date 23 September 2025) https://learn.microsoft.com/en-us/purview/retention-policies-copilot
- EUR-Lex, Regulation (EU) 2016/679 (GDPR), Art. 17 and Art. 99(2) https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- ICO, "Right to erasure" (UK GDPR guidance) https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/right-to-erasure/
- NIST, SP 800-88 Rev. 2, "Guidelines for Media Sanitization" (September 2025) https://csrc.nist.gov/pubs/sp/800/88/r2/final; full text https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r2.pdf sections 3.1.1 to 3.1.3
- Illustrative running example of this book; invented
- Regulation (EU) 2016/679, Official Journal L 119, 4 May 2016, Arts. 12 and 17, EU Publications Office copy https://publications.europa.eu/resource/cellar/3e485e15-11bd-11e6-ba9a-01aa75ed71a1

Chapter 12. Shadow Data: Finding Every Copy

- Microsoft Support, "File storage in Microsoft Teams" https://support.microsoft.com/en-us/office/file-storage-in-microsoft-teams-df5cc0a5-d1bb-414c-8870-46c6eb76686a
- Microsoft Support, "Save disk space with OneDrive Files On-Demand for Windows" https://support.microsoft.com/en-us/office/save-disk-space-with-onedrive-files-on-demand-for-windows-0e6860d3-d9f3-4971-b321-7092438fb38e
- Microsoft Learn, "XlSheetVisibility enumeration (Excel)" https://learn.microsoft.com/en-us/office/vba/api/excel.xlsheetvisibility
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" (ms.date 9 July 2026) https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy; Microsoft Learn, "Learn about retention for Copilot and AI apps" (ms.date 23 September 2025) https://learn.microsoft.com/en-us/purview/retention-policies-copilot
- Microsoft Learn, "Manage audit log retention policies" (ms.date 19 June 2026) https://learn.microsoft.com/en-us/purview/audit-log-retention-policies
- Microsoft Learn, "Learn about retention for SharePoint and OneDrive" (ms.date 22 September 2025) https://learn.microsoft.com/en-us/purview/retention-policies-sharepoint
- ICO, Penalty Notice, Police Service of Northern Ireland (26 September 2024) https://ico.org.uk/media2/migrated/4031177/psni-penalty-notice.pdf
- Review of the PSNI data breach of 8 August 2023, "Protecting From Within" https://www.nipolicingboard.org.uk/files/nipolicingboard/2023-12/review-of-the-psni-data-breach-8-august-2023.pdf publication date per the ICO penalty notice, footnote 14
- Illustrative running example of this book; invented

Chapter 13. The Assistant Reads Your Files

- Microsoft Tech Community, "Microsoft 365 Copilot is generally available" (1 November 2023) https://techcommunity.microsoft.com/blog/microsoft-copilot-blog/microsoft-365-copilot-is-generally-available/3969331
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" (ms.date 9 July 2026) https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
- Google Workspace Learning Center, "What controls Gemini's access to Workspace data" https://support.google.com/a/users/answer/17010577

- Microsoft Learn, "Considerations for Microsoft Purview to manage Microsoft 365 Copilot and Channel Agent in Teams" (ms.date 30 June 2026) https://learn.microsoft.com/en-us/purview/ai-m365-copilot-considerations
- Microsoft Learn, "Microsoft Purview DLP for Microsoft 365 Copilot and Cowork" (ms.date 9 November 2026) https://learn.microsoft.com/en-us/purview/dlp-microsoft365-copilot-location-learn-about
- Microsoft Learn, "Restrict discovery of SharePoint sites and content" (ms.date 11 October 2026) https://learn.microsoft.com/en-us/sharepoint/restricted-content-discovery
- Microsoft Learn, "Initiate site access reviews for Data access governance reports" (ms.date 9 July 2026) https://learn.microsoft.com/en-us/sharepoint/site-access-review
- Microsoft Learn, "Restricted SharePoint Search" (ms.date 6 July 2026) https://learn.microsoft.com/en-us/sharepoint/restricted-sharepoint-search
- Microsoft Learn, "Overview of Loop storage" (ms.date 14 May 2026) https://learn.microsoft.com/en-us/microsoft-365/loop/loop-storage
- Illustrative running example of this book; invented

Chapter 14. Authority and Inventory

- Illustrative running example of this book; invented
- Microsoft Learn, "OneDrive retention and deletion" https://learn.microsoft.com/en-us/sharepoint/retention-and-deletion
- Google Workspace Admin Help, "Transfer Drive files to a new owner as an admin" https://support.google.com/a/answer/1247799
- Microsoft Learn, "Create a SharePoint site ownership policy" https://learn.microsoft.com/en-us/sharepoint/create-sharepoint-site-ownership-policy
- Microsoft 365 Blog, "Announcing Microsoft 365 Copilot general availability and Microsoft 365 Chat", 21 September 2023 https://www.microsoft.com/en-us/microsoft-365/blog/2023/09/21/announcing-microsoft-365-copilot-general-availability-and-microsoft-365-chat/
- Official Microsoft Blog, "Furthering our AI ambitions: Announcing Bing Chat Enterprise and Microsoft 365 Copilot pricing", 18 July 2023 https://blogs.microsoft.com/blog/2023/07/18/furthering-our-ai-ambitions-announcing-bing-chat-enterprise-and-microsoft-365-copilot-pricing/
- Microsoft Learn, "Manage unlicensed OneDrive user accounts" https://learn.microsoft.com/en-us/sharepoint/unlicensed-onedrive-accounts

Chapter 15. Permissions, Retirement and Proof

- Illustrative running example of this book; invented
- Regulation (EU) 2016/679, Official Journal L 119, 4 May 2016 https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- Microsoft Learn, "Data access governance reports for SharePoint sites" https://learn.microsoft.com/en-us/sharepoint/data-access-governance-reports
- Microsoft Learn, "Manage sharing settings for SharePoint and OneDrive in Microsoft 365" https://learn.microsoft.com/en-us/sharepoint/turn-external-sharing-on-or-off
- Microsoft Learn, "How shareable links work in OneDrive and SharePoint in Microsoft 365" https://learn.microsoft.com/en-us/sharepoint/shareable-links-anyone-specific-people-organization
- Microsoft Learn, "Restrict discovery of SharePoint sites and content" https://learn.microsoft.com/en-us/sharepoint/restricted-content-discovery
- Microsoft Learn, "Restricted SharePoint Search" https://learn.microsoft.com/en-us/sharepoint/restricted-sharepoint-search
- Microsoft Learn, "Overview of Microsoft 365 Archive" https://learn.microsoft.com/en-us/microsoft-365/archive/archive-overview
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
- Microsoft Learn, "Manage audit log retention policies" https://learn.microsoft.com/en-us/purview/audit-log-retention-policies
- Microsoft Learn, "Use sharing auditing in the audit log" https://learn.microsoft.com/en-us/purview/audit-log-sharing

Epilogue: The Copy Outlives the Desk

- Illustrative running example of this book; invented
- Microsoft Learn, "Data, Privacy, and Security for Microsoft Copilot" https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
- Google Workspace Learning Center, "What controls Gemini's access to Workspace data" https://support.google.com/a/users/answer/17010577
- Microsoft Support, "Draft and add content with Copilot in Word" https://support.microsoft.com/en-us/word/copilot/draft-and-add-content-with-copilot-in-word

Appendix A. The Ledger for Files: Template, Inventory and Readiness Check

- Illustrative running example of this book; invented

Appendix B. Timeline

- The MS-DOS Encyclopedia, Section I (Microsoft Press, 1988) https://www.pcjs.org/documents/books/mspl13/msdos/encyclopedia/section1/
- Computer History Museum, Timeline of Computer History: Software and Languages https://www.computerhistory.org/timeline/software-languages/
- IBM History, "The IBM PC" https://www.ibm.com/history/personal-computer
- Microsoft Learn, "The History of Microsoft - 1992" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1992
- Computer History Museum, "PC Software Workshop: Lotus 1-2-3" transcript https://archive.computerhistory.org/resources/access/text/2012/04/102658156-05-01-acc.pdf
- Microsoft Learn, "The History of Microsoft - 1985" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1985
- Microsoft Learn, "The History of Microsoft - 1987" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1987
- Microsoft Learn, "The History of Microsoft - 1989" https://learn.microsoft.com/en-us/shows/history/history-of-microsoft-1989
- EUR-Lex, Directive 95/46/EC https://eur-lex.europa.eu/eli/dir/1995/46/oj
- Microsoft News, "Leading Systems Integrators Ready for the Availability Of Microsoft SharePoint Portal Server" https://news.microsoft.com/source/2001/03/28/leading-systems-integrators-ready-for-the-availability-of-microsoft-sharepoint-portal-server/
- RFC Editor, RFC 4180 https://www.rfc-editor.org/info/rfc4180
- Ecma International, ECMA-376 https://ecma-international.org/publications-and-standards/standards/ecma-376/
- Microsoft News, "Microsoft Launches Office 365 Globally" https://news.microsoft.com/source/2011/06/28/microsoft-launches-office-365-globally/
- Official Google Blog, "Introducing Google Drive… yes, really" https://googleblog.blogspot.com/2012/04/introducing-google-drive-yes-really.html
- JPMorgan Chase Form 8-K https://www.sec.gov/Archives/edgar/data/19617/000001961713000077/0000019617-13-000077.txt
- Microsoft News, "Microsoft OneDrive launches worldwide" https://news.microsoft.com/source/2014/02/19/microsoft-onedrive-launches-worldwide-free-cloud-storage-service-provides-one-place-for-everything-in-your-life/
- https://doi.org/10.1186/s13059-016-1044-7
- Regulation (EU) 2016/679 https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- GOV.UK, "PHE statement on delayed reporting of COVID-19 cases" https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases
- ICO, Penalty Notice: Police Service of Northern Ireland https://ico.org.uk/media2/migrated/4031177/psni-penalty-notice.pdf
- Microsoft Tech Community, "Microsoft 365 Copilot is generally available" (1 November 2023) https://techcommunity.microsoft.com/blog/microsoft-copilot-blog/microsoft-365-copilot-is-generally-available/3969331
- NIST CSRC https://csrc.nist.gov/pubs/sp/800/88/r2/final
- Federal Reserve, SR 26-2 https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm

- Microsoft Learn, "Restricted SharePoint Search" https://learn.microsoft.com/en-us/sharepoint/restricted-sharepoint-search

Appendix C. Failure Library

- Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses https://elischolar.library.yale.edu/ypfs-documents/454; Form 8-K https://www.sec.gov/Archives/edgar/data/19617/000001961713000077/0000019617-13-000077.txt
- PERI, University of Massachusetts Amherst https://peri.umass.edu/publication/item/526-does-high-public-debt-consistently-stifle-economic-growth-a-critique-of-reinhart-and-rogoff
- Genome Biology 17:177 (2016) https://doi.org/10.1186/s13059-016-1044-7; https://pmc.ncbi.nlm.nih.gov/articles/PMC4994289/
- GOV.UK https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases
- ICO, Penalty Notice: Police Service of Northern Ireland https://ico.org.uk/media2/migrated/4031177/psni-penalty-notice.pdf; enforcement page dated 3 October 2024 https://ico.org.uk/action-weve-taken/enforcement/2024/10/police-service-of-northern-ireland-mpn/

THE SERIES

Every Record Counts

Every Record Counts is a series about data: where each record was born, what it means, who can change it, what it costs and who reads it now, from the tally stick to the training set. Each book stands on its own.

Also by the author:

Every Record Counts	**From Tally to Token**: the record, its readers, and what each era broke
Every Record Counts	**The Model as Reader**: data for training, retrieval and agents
Every Record Counts	**From Warehouse to Lakehouse**: data on every platform, and the price of exit
Every Record Counts	**What a Value Means**: types, encodings and formats
Every Pixel Matters	**From Bit to Pixel**: the whole stack, from the bit to the pixel
Every Pixel Matters	**From Code to Glass**: the rendering layer
Every Pixel Matters	**From Silicon to Screen**: who controls each layer, and what trust costs
Medallion	**From Bronze to Gold**: why data layers decide what AI can do

Juan Cardena

Enterprise AI & Agentic Systems Architect · blog.jcardena.com

Download the latest edition and the blank Ledger at blog.jcardena.com/en/ebooks/data-on-every-desk.

Every Record Counts · Book 4 · Data on every desk.