Privacy **Guide**

How privacy works in a DataSupply deal: who touches what, how personal information is removed or transformed, and what stays in the owner's control.

AI labs want how the work gets done, not who your customers are. They pay for real workflows and real conversations. Names, emails and numbers are removed or replaced before the data is used.

Never	You	Before use
RAW DATA NEVER PASSES THROUGH DATASUPPLY	THE COMPANY DECIDES WHAT IS IN AND WHAT IS OUT	THE AI LAB DE-IDENTIFIES THE DATA UNDER THE LICENSE TERMS

STEP BY STEP

Who touches what, and when

	STEP	WHAT HAPPENS	RAW DATA MOVES?
01	**Discovery meeting**	A conversation about the company's systems and goals. No files, exports or logins. DataSupply does not sign an NDA at this stage because no confidential data is shared.	NO
02	**Agreement**	The company signs the Data Commercialization Agreement with DataSupply. It covers how DataSupply represents the data, not delivery of data.	NO
03	**Manifest**	The company describes its data: which systems, how much, how many years. DataSupply only collects this information about the data, never the data itself.	NO
04	**Anonymized profile**	DataSupply presents the manifest to vetted AI labs as an anonymized profile. The company's name is shared only with a buyer it approves.	NO
05	**Offers and approval**	Labs compete with offers. The company approves the buyer, the scope and every term in writing, and decides what is excluded.	NO
06	**Secure delivery**	The company pushes only the approved, scoped export straight to the AI lab through an encrypted, time limited transfer. Nobody logs into the company's systems.	COMPANY TO LAB
07	**De-identification**	Inside its secure environment, the AI lab removes or transforms personal information under the license terms, before the data is used. Software detects personal information and people review samples.	STAYS AT LAB
08	**License term and end**	The lab uses the data only as the license allows. When the license ends, its deletion or return terms apply.	NO

DataSupply never receives, stores, hosts or sees the raw data at any step. Partners never collect data, files or passwords.

Two ways to protect people: redaction and transformation

De-identification uses two families of methods. Most deals use both: redact what the lab does not need, and transform what it needs so the workflow still makes sense.

REDACTION

Take it out

Removing or withholding data entirely. The detail is gone, not hidden.

BEFORE
Anna emailed John about Acme Corp.

AFTER
[REDACTED] emailed [REDACTED] about [REDACTED].

TRANSFORMATION

Change it, keep it useful

Swapping identifying details for consistent stand ins, so the data stays useful but no one is identifiable.

BEFORE
Anna emailed John about Acme Corp.

AFTER
John emailed Sam about ABC Company.

METHOD	BEFORE	AFTER
REDACTION: REMOVING OR WITHHOLDING		
Drop fields or columns	Customer table with SSN column	SSN column deleted
Exclude records or systems	Email, CRM, HR files	Email, CRM (HR files never exported)
Suppression	Note: "Call Mike Ross re: invoice 88412"	Note: "Call [REDACTED] re: [REDACTED]"
Minimization	Every folder, all years	Only the systems and years the license needs
TRANSFORMATION: CHANGING SO IT IS NO LONGER IDENTIFYING		
Pseudonymization	Anna Lee, anna@acme.com	Person_0412, person_0412@example.com (same stand in every time)
Masking	Phone 604 555 0199	Phone XXX XXX 0199
Generalization	Age 37; 1420 Oak Street, Vancouver	Age 35 to 44; Western Canada
Date shifting	Opened Mar 3, closed Mar 10	Opened May 14, closed May 21 (order and gaps kept)
Aggregation	Each invoice by customer	Monthly totals by region
Synthetic replacement	Real customer names in a ticket thread	Realistic made up names, same story

When each is used

- **Redaction** for anything the lab does not need or that must never leave: government ID numbers, card and bank numbers, passwords, health details, whole systems the owner excludes.
- **Transformation** for the people and companies inside the work: names, emails, phone numbers, addresses and dates, so threads, tickets and records still connect.

The tradeoff

Redaction — Transformation

Redaction protects by removing, but it lowers value the most: a thread full of blanks loses its story. Transformation keeps workflows and connections intact, so it keeps more value. The company decides what is excluded; labs value what is left.

Stand in keys that link a real name to its replacement are never delivered with the data.

CONTROL AND PROTECTION

What the owner controls, what the lab must do, what never goes

The owner controls

- Which systems are in, and which are opted out.
- Which fields, folders, records or years are excluded, before anything is exported.
- Which buyer, which scope and which terms. Nothing moves without written approval.
- Samples: reviewed and approved before delivery.
- A rights review of its own privacy policy, customer contracts and staff policies, so it only licenses what it has the right to license. Its own counsel is welcome to review.
- Whether to say no to any offer.

Each license requires

- The lab to de-identify the data before it is used.
- Use it only for the permitted uses, such as training AI models.
- Never try to re-identify anyone or reverse engineer the business, and never resell it without consent.
- Keep it isolated and encrypted, with time limited, revocable access and prompt notice of any incident.
- Follow deletion or return terms written into the license. Whether a lab also gives a separate written deletion confirmation depends on the lab and is agreed per deal.

Never licensed

- Data involving minors.
- Export controlled or classified information.
- Passwords and credentials.
- Card and bank account numbers.
- Health and biometric data.
- Privileged legal material.
- Content owned by clients or third parties, or covered by a contract that forbids sharing.

> **Regulations.** Privacy rules depend on the country and the industry. Each license sets out its own privacy and security terms, and DataSupply pauses if a rights or privacy question comes up until it is resolved. DataSupply does not give legal advice, and no deal is described as compliant or certified under any specific law. The company's own counsel should review before it signs.

QUICK ANSWERS

Will my customers' names go to an AI lab?	Not in a form anyone can use to identify them. The lab removes or replaces names, emails and phone numbers before the data is used, and the license bans any attempt to re-identify people. Sources you would rather keep private can be left out entirely.
Do you see our data?	No. DataSupply only collects information about your data in the manifest: which systems, how much and how long. Raw data goes straight from you to the AI lab and never passes through DataSupply.
Can we exclude something?	Yes. You can opt out whole systems, specific fields, folders, people or years. Exclusions are agreed before delivery and only the approved export leaves.
What happens after the license ends?	The license's deletion or return terms apply to the data the lab received. Models already trained usually stay in use, because data cannot be pulled back out of a trained model, which is why scope is settled before delivery.