



GHOST
PROTOCOL

OS

THE FIELD LIBRARY

OPERATIONS

LOOK BEFORE YOU TOUCH

OSINT and Recon

The recon mindset, infrastructure and people OSINT, breach data, and the attack surface map.



Ghost Protocol (Pvt) Ltd · Colombo · ghosts.lk

Security internship training · The Field Library · Edition 1 · 2026

THE GHOST PROTOCOL FIELD LIBRARY · OPERATIONS

OSINT and Recon

The recon mindset, infrastructure and people
OSINT, breach data, and the attack surface map.

Ghost Protocol (Pvt) Ltd · Colombo

The Field Library, Edition 1 · 2026

Copyright 2026 Ghost Protocol (Private) Limited. All rights reserved.

Sold as a PDF for the buyer's own reading. Please do not share or resell the file.

The library

The Field Library is a whole shelf, not one book. The Core Path is the spine everyone reads. The rest is depth and breadth, taken on as the work calls for it. The shelf is complete now, and you are reading the one marked here.

A · THE CORE PATH

O The Oath BUILT

I Groundwork BUILT

II The Web BUILT

III The Wider Field BUILT

IV The Work BUILT

V The Craft BUILT

B · DEEP TRACKS

AI and Agents BUILT

Reading Code BUILT

Internal and Active Directory BUILT

The Web, Deeper BUILT

Cloud and Containers BUILT

Mobile and APIs BUILT

Reverse Engineering BUILT

Hardware and the Edges BUILT

C · THE OPERATIONS

Red and Purple Team BUILT

OSINT and Recon YOU ARE HERE

Report Writing Mastery BUILT

The Consulting Craft BUILT

Bug Bounty BUILT

D · THE OTHER SIDE

Detection and Response BUILT

Threat Hunting and Forensics BUILT

Hardening and Architecture BUILT

E · THE GROUND RULES

The Law, Deeper BUILT

Standards and Compliance BUILT

Ethics and Disclosure BUILT

F · COMPANIONS

The Reference Pack BUILT

The Lab Workbook BUILT

Case Files BUILT

The Bestiary BUILT

The Glossary and the Ledger BUILT

Contents

- Opening 5
- 1** The Recon Mindset 6
- 2** Infrastructure OSINT 11
- 3** People and Organisation OSINT 16
- 4** Breach Data and Exposure 20
- 5** The Attack Surface Map 24
- Worked answers 29

Look before you touch



The engagement had just started and Blink's cursor was already hovering over the target, keen to send the first packet.

BLINK "Can we scan it yet. Can we. Can we."

SOL "Not yet. Before you touch the target you learn everything the open world already knows about it: its domains, its servers, the addresses it forgot it owned, the names on the records. All of it without a single packet the target would ever see. The quietest hour of the engagement teaches you the most, and it teaches the defender nothing."

Worm's eye was already on a certificate log, reading names the client did not know were public.

WORM "look. don't touch."

And one line runs under the whole book: public is not the same as fair. You will map an organisation's surface, not dig through a person's private life, and the difference between those two is the difference between recon and harm.

CHAPTER ONE

The Recon Mindset

Before you touch a target you learn it, and you learn it from everything except the target itself. This chapter sets the order of the work, the footprint you build, and the line recon never crosses.

Most of a test is decided before you send the first request to the target. Recon is the quiet half of the work: hours spent reading what the world already says about a system, so that when you finally engage it you are aiming, not wandering. A tester who skips it finds a fraction of the surface and spends the rest of the engagement surprised. This volume is how you stop being surprised.

Volume 0 set the ground you stand on: you test only what you are authorised, in writing, to test, and personal data you meet is evidence, not a trophy. Volume II gave the test its shape, and named its second movement Map, the step where you learn an application as a system before you probe it. This volume is that movement opened out to its full size. Recon is Map done properly, and done first, across infrastructure, people, and exposure, for a target you have not yet lawfully touched.

Open-source intelligence, defined narrowly

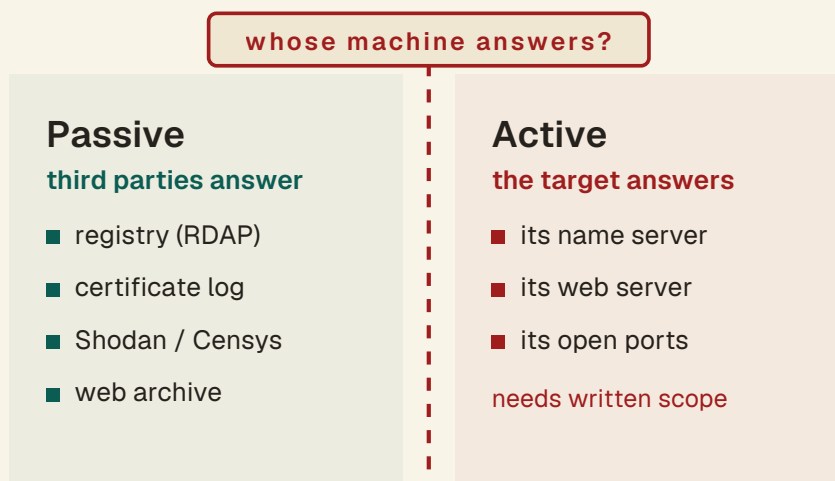
Open-source intelligence, OSINT, is the practice of building a picture of a target from information that is already public: the domain records, the certificates, the archived pages, the code a company pushed to a public repository, the job adverts that name its internal tools. None of it is secret, and none of it requires you to break in to read. It requires you to know where to look and to read carefully. The skill is not access. The skill is assembly: a hundred ordinary public facts, put next to each other, become a map that no single source meant to draw.

Hold the word public strictly. A thing is public if anyone may read it without defeating a control, without a password you were not given, without a term of service you agree to break. A login page is public; the account behind it is not. A company's team page is public; the mailbox it lists is not a door you are invited through. OSINT lives entirely on the near side of that boundary, and the moment a technique needs you to cross it, you have left OSINT and entered testing, which needs the paper Volume 0 described.

Passive before active

There is an order to recon, and it is the same order Volume II drilled for the web. Passive comes before active, and the dividing line is simple to state: passive work never sends anything to the target's own systems. You read third parties and the public record, the registries and the search indexes and the archives, all of which hold information about the target without being the target. Active work is anything the target's own machine answers: resolving its hostnames against its own name server, fetching its pages, scanning its ports. Active work belongs to the engagement, under scope, and this volume hands most of it to Volume II and Volume III.

The distinction is not academic. It decides whether you are inside your authorisation or outside it. A certificate transparency search reads a public log that certificate authorities publish; you could run it against a company you have never met and break no rule, because you never touched them. A port scan of the same company, with no written scope, can be charged as unauthorised access under the Computer Crime Act, No. 24 of 2007, and we treat it as off-limits, exactly as Volume 0 set out. Same curiosity, two different sides of a line, and the line is whose machine answered the question.



Passive recon reads third parties and never touches the target; the moment the target's own machine answers, you are active and need scope.

■ THE LINE

Recon has two boundaries, not one, and you check both before every lookup.

- **Whose machine answers?** If the target's own system responds to your query, you are doing active testing and you need written scope that names it. Passive OSINT only ever touches third parties and the public record.
- **Does the engagement need it?** Public does not mean collectible. You gather only what a specific authorised engagement actually requires. Everything else, however easy to reach, you leave where it lies.

A yes to the first question without written scope is a stop. A no to the second is a stop even when the data is lying open. When unsure, you ask before you look, never after.

The Footprint you build, and the one you leave

Recon builds a footprint of the target: the set of domains, hosts, addresses, services, people, and public exposures that together describe how much of the organisation faces the world. That footprint is the product of this volume, and the next four chapters each add a layer to it. But you also leave a footprint of your own while you work, and a professional keeps that in mind from the first lookup.

Every tool sits on one side of the passive line, and you should be able to say which before you run it. A WHOIS or RDAP query hits a registry, not the target. A certificate transparency search hits a public log aggregator. A search of archived pages hits Archive.org. These leave no trace on the target at all.

Resolving a hostname, on the other hand, usually reaches the target's authoritative name server; fetching a page reaches its web server; and a scanner reaches everything. Knowing which category a tool falls into is not pedantry, it is the difference between quiet recon and an unauthorised touch you did not mean to make.



whose?

Before a lookup, one question answers both boundaries at once: whose machine answers this, and does the engagement need the answer. If you cannot say, you have not earned the keystroke yet.

Why recon decides the test

A scope names a target, but a scope is not a plan. The plan is what recon produces. When you know the real size of an organisation's surface, which hosts are live, what they run, where the old forgotten box sits, you spend your authorised hours on the parts most likely to break, and you do not waste a day on the three pages the client expected you to look at while the vulnerable admin panel sat on a subdomain nobody mentioned. Half of what separates a thorough test from a shallow one is simply knowing the surface exists. You cannot test what you never found.

Recon also protects the scope. The surface map tells you not only what is in bounds but what sits just outside it, the neighbouring host that belongs to the same company but is not on your paper. Finding it in recon means you record it and leave it alone deliberately, rather than stumbling onto it mid-test and reaching one hop too far. The map is how you stay inside the lines while still seeing the whole field.

Output is a lead, not a fact

Public sources are stale, incomplete, and sometimes wrong. An archived page shows what was true in 2023. A search result repeats a claim nobody checked. A breach-notification hit tells you an address appeared in some corpus, not that the password still works or ever did. Treat every piece of recon as a lead that points somewhere, not a fact you can state. The verification discipline from Volume II applies here in full: before anything built on recon goes into a report, it is confirmed against current, primary evidence, inside scope. The map tells you where to look. It does not tell you what is true.



SOL SAYS

On the first day of an engagement the beginner wants to run something, because running something feels like progress. The professional reads. They pull the public record, they draw the surface, they sit with it, and from the outside they look like they are doing nothing at all. They are doing the thing that decides everything after it. The hours you spend knowing the target are the hours that make the rest of the test fast, and they are free, and they are quiet, and they are yours to take before anyone can say you touched a thing you should not have.

The privacy line, stated once

Two of the chapters ahead touch real people: the staff of an organisation, their names and roles and work addresses. People-OSINT is lawful when it is relevant and restrained, and it turns ugly the instant

it is neither. The whole of the discipline is a single filter, applied before you collect anything about a person: is this about the engagement, or about the person? An email naming convention that lets you reason about a company's logins is about the engagement. A home address, a family photo, a private account, a movement pattern, these are about the person, and they do not belong in your notes no matter how public the page that held them. Sri Lanka's Personal Data Protection Act, No. 9 of 2022, which Volume 0 covers, raises the stakes on getting this wrong. It does not change what right looks like, which was always the same: you take only what proves the engagement, and never the thing that only serves curiosity or pressure. Clean conduct, from Volume 0, is not suspended because the data was easy to find.

■ EXERCISE

Pick a target you are allowed to recon: a domain you own, or `ghosts.lk`, which as the house's own domain you may recon from passive sources only; anything active against it needs Ryan's written go-ahead first. Before running anything, write a two-column list. In the left column, name ten lookups you might make in the chapters ahead (a WHOIS query, a certificate transparency search, a DNS resolution, a page fetch, a port scan, an archived-page read, a GitHub search, and so on). In the right column, for each one, write passive or active, and name whose machine answers it. The deliverable is the list, not the results. You are proving you can tell, before you act, which side of the line a keystroke lands on.

■ VERIFY IT

Take one claim from any public source about your chosen target, something as small as "this domain was registered in 2019" or "this host runs nginx", and prove it two independent ways before you believe it. Two sources that trace back to the same original are one source wearing two coats, so look for genuine independence: a registry record and an archived page, a certificate log and a header you are authorised to read. If the two disagree, you have learned the most useful thing recon teaches, which is how often the easy answer is wrong.

■ ON YOUR FEET

You are scoped to test a single web application for a client. An hour into passive recon, an ordinary search turns up the personal blog of the client's chief executive: their home town, photographs of their family, the name of their child's school, all posted publicly by the executive themselves over several years. None of it is a security finding and none of it is in your scope. It is sitting there, fully public, one click away. What do you do with it, and what do you not do?



WHAT WYRM KEEPS

Recon is the half of the test you do before you touch the target, from everything except the target. Passive before active. Public does not mean collectible. A lead is not a fact.

- The passive line is whose machine answers: third parties and the public record are passive, the target's own systems are active and need written scope.
- Recon decides where the authorised hours go and keeps you inside the lines by showing you where they are.
- Collect only what the engagement needs. People-OSINT is filtered by one question: is this about the engagement, or about the person? The PDPA and clean conduct (Volume 0) are not suspended because the data was easy to find.
- Every recon result is a lead, verified against primary evidence before it becomes a claim.