

Lab Guide

Five exercises: WHOIS analysis, DNS subdomain mapping, certificate transparency mining, social profile correlation, and paste-dump triage — all AI-assisted.

Authorized Use Only

All challenge files are **synthetic**. Do not run lookups against the fictitious domains in these files — they may resolve to real IP space. The techniques you learn here are only for authorized engagements.

Prerequisites

- Completed the [Theory](#) reading
- Access to an AI assistant (Claude, ChatGPT, or local LLM via Ollama)
- Downloaded all five files from the [Theory page resources](#) into a local folder named `osint-lab/`
- A text editor and (optional) `jq`, `grep`, `python3`

Lab Scenario

You have been hired by **AcmeCorp** to perform an authorized red team engagement. Your scope document lists the domain `acmecorp.com` as in-scope. Your goal in this lab is to build a complete reconnaissance picture of AcmeCorp using only the OSINT files provided — identifying their infrastructure, employees, exposed credentials, and the most promising initial-access vectors.

Download Lab Files

Save all five files into a folder named `osint-lab/` before starting the exercises.

↓ [whois-records.txt](#) ~5 KB

↓ [dns-records.txt](#) ~4 KB

↓ [certificate-transparency.txt](#) ~6 KB

↓ [social-profiles.txt](#) ~8 KB

↓ [paste-dump.txt](#) ~5 KB

Exercise 1 — WHOIS Triage

Open [whois-records.txt](#). The file contains WHOIS records for several domains.

Task 1.1 — Identify the target's primary domain

1. Read through [whois-records.txt](#) manually for 2–3 minutes. Note any patterns you see in registrant info, name servers, and dates.
2. Paste the entire file into your AI assistant with this prompt:

You are a red team analyst. Below is a WHOIS dump with multiple domains. For each domain, extract:

- Registrar
- Creation date and expiry date
- Registrant organization (if visible)
- Name servers
- Any technical or admin contact emails
- Privacy protection status (yes/no)

Then identify which domains likely belong to the same organization based on name servers, registrant patterns, or contact emails. Output as a markdown table.

[paste whois-records.txt here]

Compare the AI's output to your manual reading. Did the AI catch anything you missed?

✓ **Self-check:** You should be able to name (a) which domains share infrastructure with `acmecorp.com`, and (b) at least one contact email that could be used as a phishing lure or username guess.

Exercise 2 — DNS & Subdomain Mapping

Open `dns-records.txt`. This file contains DNS records (A, AAAA, MX, TXT, CNAME, NS) for AcmeCorp's domains.

Task 2.1 — Build the attack-surface map

Ask the AI to convert the DNS dump into an attack-surface inventory:

Below is a DNS record dump. Build an attack-surface inventory:

1. List every unique hostname.
2. Group hostnames by service type (web, mail, VPN, dev/staging, admin, ...).
3. Flag any hostname that suggests it should NOT be internet-facing (admin panels, staging environments, internal apps, debug endpoints).
4. For each MX record, note the mail provider.
5. For each TXT record, identify SPF/DKIM/DMARC posture and any leaked third-party services (e.g., Salesforce, Atlassian, Office365).

Output:

- Section 1: Inventory table
- Section 2: Top 5 highest-priority targets and why

[paste dns-records.txt]

✓ **Self-check:** You should have a ranked list of subdomains and at least 2 candidates for a "low-hanging fruit" initial access target.

Task 2.2 — Cross-reference with WHOIS

Paste both the WHOIS table from Exercise 1 and the DNS inventory from Task 2.1 into the AI. Ask:

"Which subdomains are hosted on AcmeCorp's own IP space versus a third-party provider? Which findings correlate across the two datasets, and which appear only in one?"

This is the kind of correlation that takes a human analyst 30 minutes and an AI assistant 5 seconds.

Exercise 3 — Certificate Transparency Mining

Open `certificate-transparency.txt`. CT logs reveal subdomains that may not appear in DNS yet, or that the organization tried to hide.

Task 3.1 — Find the hidden assets

Below is a Certificate Transparency log extract. Extract:

1. Every unique subdomain (Subject Alternative Names included).
2. Issue date and certificate authority for each.
3. Any wildcard certificates (and what they cover).
4. Any subdomain that suggests internal/staging use (dev-, staging-, qa-, internal-, admin-, vpn-, test-).
5. Any certificates issued in the last 90 days (recent activity = active)

Then compare against this list of subdomains I already know about:

[paste your subdomain list from Exercise 2]

Output the DELTA: subdomains found in CT logs that are NOT in my known list

✓ Self-check: The CT delta should reveal at least one subdomain that did not appear in `dns-records.txt`. That subdomain is your highest-priority lead.

Exercise 4 — Social & People Recon

Open `social-profiles.txt`. This contains LinkedIn-style employee profiles for AcmeCorp staff.

Task 4.1 — Build the org chart

Below are employee profiles for AcmeCorp. Extract:

1. Name, title, department, location for each person.
2. Build an org chart hierarchy (CEO → VPs → directors → managers → ICs).
3. Flag anyone in IT, Security, DevOps, or Engineering — they have the most valuable access.
4. Flag anyone whose bio mentions specific technologies, tools, or internal projects (these are reconnaissance gold).
5. For each "high-value" person, suggest 2-3 plausible username formats given the company's email pattern (you may have to infer the pattern).

[paste social-profiles.txt]

✓ **Self-check: You should have at least 3 high-value targets, each with 2–3 username candidates ready for password-spray or phishing.**

Task 4.2 — Pretext generation (defensive learning)

Ask the AI:

"For the highest-value target identified above, draft three plausible phishing pretexts that a real attacker would use. For each, explain what defensive control would block it. This is for security-awareness training."

Notice how the AI's framing changes the answer quality. A red team practitioner uses this to test their client's defenses; defenders use the same exercise to identify gaps.

Exercise 5 — Paste-Dump Triage

Open `paste-dump.txt`. This is a collection of pastes scraped from public paste sites that mention AcmeCorp.

Task 5.1 — Find leaked credentials and secrets

Below are public paste-site dumps mentioning AcmeCorp. Extract:

1. All credentials (username:password, email:password, API keys, tokens).
2. Any internal hostnames, IP addresses, or URLs disclosed.
3. Any source-code snippets and what they reveal (DB connection strings, hard-coded secrets, internal endpoints).
4. Any references to internal projects, codenames, or employee names.
5. The estimated dump date based on file metadata or content.

Rank the findings by exploit value: which leak gives the most direct path to initial access?

```
[paste paste-dump.txt]
```

✓ **Self-check:** You should be able to point to at least one credential or token that could be used *right now* against an asset identified in earlier exercises.

Final Exercise — The Recon Report

Task 6.1 — Synthesize everything

Open a new conversation with your AI assistant. Paste the outputs from Exercises 1–5 and ask:

Synthesize these five OSINT outputs into a single red team reconnaissance report. The report must contain:

1. Executive Summary (3 sentences max)
2. Confirmed Infrastructure (domains, subdomains, mail, hosting)
3. People of Interest (top 3, with username candidates)
4. Exposed Credentials and Secrets
5. Top 3 Initial Access Recommendations, ranked by:
 - Likelihood of success

- Stealth (will it trip alarms?)
- Time-to-exploit

6. Open questions / gaps that require active scanning to resolve

Format as markdown. No more than 800 words total.

✓ **Self-check: The final report should clearly identify your #1 attack path and the specific evidence supporting it.**

Reflection

- How much faster was the AI-assisted workflow compared to manually grepping each file?
- What did the AI miss that you caught? (There is always something.)
- Which OSINT source had the highest "signal-to-noise" ratio for this target?
- If you were the defender, which finding would you remediate first?

What's Next

Lesson 02 takes you from passive recon to **log analysis** — using AI to read attacker activity from the inside, including authentication logs, web access logs, and Windows event logs.