

# Open Source for Cyber Defence/Progress

From GCWiki  
 < [Open Source for Cyber Defence](#)  
 Jump to: [navigation](#), [search](#)

Many structured datasets are now available in the [HAPPY TRIGGER](#) database. Unstructured datasets are being worked on and will go to [LOVELY HORSE](#). Other integration with [TWO FACE](#) and [Zeol](#), is in place, and more will come to [XKEYSCORE](#).

## Contents

- 1 [Data currently gathered](#)
- 2 [Future ones to work on](#)
  - 2.1 [Vulnerability Intelligence](#)
  - 2.2 [Bulk Infrastructure Data](#)
  - 2.3 [Miscellaneous](#)

### [edit] Data currently gathered

| Data source                          | Nature of the data  | OPP-LEG Status                        | In HAPPY TRIGGER?              | In LOVELY HORSE?         | In Zeol?                 | In TWO FACE?             | Update frequency   |
|--------------------------------------|---|---------------------------------------|--------------------------------|--------------------------|--------------------------|--------------------------|--|
| alexa.com                            | Top domains list; has previously been used to find popular social networking sites in foreign countries to help with analyst investigations.  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Automatic updates on daily basis   |
| user-agents.org                      | User agent strings, useful for finding spoofed or malicious entries   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Manual update  |
| www.nsrlist.gov                      | Access to hashes of known COTS files  | Approved (for free scrape)            | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Manual update every three months   |
| www.maxmind.com (ASN list)           | Used to help map out IP ranges of networks being monitored.   | Approved (for free scrape)            | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Manual update on best endeavours basis   |
| ZeusTracker.abuse.ch                 | Zeus specific malware tracking including IPs, binaries and domains to be used by the e-crime team.  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Automatic updates on hourly basis  |
| SpyEyeTracker.abuse.ch               | SpyEye specific malware tracking including IPs, binaries and domains to be used by the e-crime team.  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Automatic updates on hourly basis  |
| amada.abuse.ch                       | Useful for declassifying information about known malicious IPs and domains.   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Automatic updates on hourly basis  |
| http://torstatus.blutmagie.de/       | TOR consensus document, useful for identifying whether a target was using TOR and the status of the individual nodes.   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Automatic updates on hourly basis  |
| EmergingThreats.net                  | Snort rules used for network monitoring purposes  | Approved (for Free data)              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Manual updates on best endeavours basis  |
| PremiumDrops.com                     | Daily newly registered domains to alert analysts to suspicious domains worth investigating for malicious activity   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Currently unavailable, need to find covert access method for paid content          |
| verisign.com                         | Monthly updates of newly registered domains to alert analysts to suspicious domains worth investigating for malicious activity  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |  |
| MalwareDomainList.com                | General malware tracking resource   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Currently one-off sample   |
| twitter.com                          | Real-time alerting to new security issues reported by known security professionals, or planned activity by hacking groups e.g. Anonymous. For more information about the sources currently being brought into the building see source list on <a href="#">the LOVELY HORSE wiki</a> | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Prototype currently running. For more information see <a href="#">LOVELY HORSE</a> |
| ContagioMiniDump.com                 | Most recommended blog by CDO analysts. Highly regarded for malware analysis relevant to APT investigations. Can be useful to declassify information for reporting purposes  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |  |
| metasploit.com                       | Access to new zero-day exploits for the malware team to analyse   | Approved (for free data)              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |  |
| exploit-db.com                       | Access to an archive of exploits and vulnerable software. Exploits from submittals and mailing lists collected into one database.   | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |  |
| ics.sans.edu (Internet Storm Center) | Already used by GovCertUK on a daily basis for timely and relevant security news and incident reporting.  | Approved                              | <input type="checkbox"/>       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Currently updated on best endeavours basis   |
| <a href="#">POSITIVE PONY</a>        | IP address to company and sector mapping. See the POSITIVE PONY wiki page for more details.   | Approved<br>Further approvals pending | <input type="checkbox"/> (dev) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Currently a static data set  |
| <a href="#">NETPLATE</a>             | Multiple data types - details will be included on this page when releasable   |                                       |                                |                          |                          |                          |  |

([POSITIVE PONY screenshots](#))

### [edit] Future ones to work on

| Knowledge required                 | Available from  | Update frequency                          | Filtering  | Volumetrics              | Comments  |
|------------------------------------|---|---|------------|--------------------------|---|
| <b>whois records</b>               | From the Passive Sigtint system, or buy from RIRs (Regional Internet Registries)? Or can we find another way of getting all updates copied to us? What about NSA's FOXTRAIL? Or our own GeoFusion? And there's now REFRIED CHICKEN from [REDACTED] ("It's a database of passively intercepted domain WHOIS records, searchable by any word in the record. Since Feb 2011. There are legal and policy constraints which mean you cannot search domains, or terms within records, that may be sensitive on grounds of location or nationality without appropriate authorisation. If you would like an account please let me know. Access to the data relies on having a Global Surge Account.") | every few days                            | don't know | don't know: ask the NAC? | NSA's <a href="#">FOXTRAIL</a> , is in this space, and needs more checks to see whether it isn't suitable. And GeoFusion (poc: [REDACTED]). |
| <b>recent domain registrations</b> | maybe an analytic run against the main DNS records to find the new domains -- or is there a more definitive source?<br>Companies like Cyveillance are able to obtain feeds of new domain registrations (for 'brand monitoring', so I imagine we'd be able to get hold of something similar... [REDACTED]@gchq 09:51, 7 September 2011 (BST))  | ready for morning and afternoon 'shifts'? | none?      | very small (MB)          | NSA's <a href="#">FOXTRAIL</a> , is in this space, and needs more checks to see whether it isn't suitable                                   |

| Site                                     | Type of data  | Legal status  |
|--|---|---|
| Pastebin                                 | An increasing number of tip-offs are coming from the Pastebin website, as this is where many hackers anonymously advertise and promote their exploits, by publishing stolen information. An automated, regular search (say, weekly) across Pastebin for certain keywords such as .gov.uk or GSI or HMG etc. would be very valuable to ensure that GovCertUK is always notified if any information that they need to be concerned about appears in open source. "30-11-2011 GovCertUK briefed about an attack on a UN server. This tip came from open source and specifically from Pastebin where the stolen emails and passwords had been posted online." | NOT APPROVED: This nature of this site means that it would be very difficult to demonstrate the proportionality of scraping the whole site to identify the small proportion of information that would be of value to CDO and therefore approval cannot be given for scraping of the site. |
| OVAL List                                | for NDR to feed into <a href="#">HIDDEN SPOTLIGHT</a> vulnerability database  | APPROVED  |
| Afrad.org                                | [REDACTED]: This lists domains which are publically available for anyone to add a sub-domain to. CDO analysts have suggested that this should be another resource they check alongside whois and robtex when investigating a domain.  |   |
| Joe Stewart's blog for Dell Secure Works | [REDACTED]: this regularly includes SNORT rules and other information that can be signatured.   | APPROVED  |
| scadasec mailing list                    | [REDACTED] request  | APPROVED  |

### [edit] Vulnerability Intelligence

| Knowledge required   | Available from   | Update frequency | Filtering   | Volumetrics     | Comments  |
|--|--|------------------|---|-----------------|---|
| <b>twitter traffic for vulnerabilities</b>                                   | use twitter API in standard way  | hourly?          | by twitter names of known malware/vulnerability researchers | very small (MB) | Current work is <a href="#">BIRD SEED</a> . JTRIG's BIRDSTRIKE provides the scraping already, but only for handfuls of IDs, and doesn't repeat. The tweets requires data mining. Experiment run by CDT for NDR using Cyber Cloud, and has OPP-LEG approval already.   |
| <b>certain blogs and CERT web sites for vulnerabilities</b>                  | direct web scrape (if allowed). <a href="#">MHS OSINT pages</a> have examples? | hourly?          | by list of specific sites/pages                             | small (GB)      | <a href="#">TR-CISA</a> have previously run several contracts looking at this problem, with a view to delivery to <a href="#">CNE</a> . Final wrap up work is scheduled to automate the derivation of SEM rules (see <a href="#">TR-FSD</a> ) from open source information such that machines matching those rule (vulnerabilities) can be found in passive. Wanted by NDR (ref <a href="#">MARBLE POLLS</a> ) and GovCERT. See <a href="#">Open source vulnerability sources</a> . |
| <b>certain CERT IRC chatrooms for vulnerabilities</b>                        | direct IRC access (if allowed)   | hourly?          | by list of specific IRCs                                    | v.small (MB)    | NB: Assume will include some encrypted IRCs. Wanted by GovCERT. Maybe a <a href="#">MARBLE POLLS</a> source.  |
| <b>certain CERT email lists for vulnerabilities</b>                          | direct reception   | hourly?          | by list of specific mailing lists                           | v.small (MB)    | NB: Assume will include some encrypted email (including PGP). Wanted by GovCERT. Maybe a <a href="#">MARBLE POLLS</a> source.   |
| <b>Commits to open source code repositories and security patch check-ins</b> | GitHub etc.  | daily?           | by specific code projects, presumably                       | small (GB)      | Requested by NDR [REDACTED].  |
| <b>Emerging Threats 'Open'</b>   | Scraped via SHORFALL framework   | Daily?           | By updated Snort rules                                      | ???             | Approval granted from OP-LEG to scrape info.  |

### [edit] Bulk Infrastructure Data

| Knowledge required                                | Available from  | Update frequency    | Filtering | Volumetrics     | Comments   |
|---|---|---------------------|-----------|-----------------|--|
| <b>known malware/bot/spam servers/orbs/relays</b> | eg, SpamHaus block lists, DNS block lists (dnsbl.abuse.ch), DNS blackholing lists (malwaredomainlist.com), Drive-by downloads (blade-defender.org) etc. | several times a day | none      | small (GB)      | SpamHaus import is already an exploit-level service from ITServices. <a href="#">TR-CISA</a> have just completed an initial study of open sources of this sort of information, with an initial delivery of sample data to <a href="#">CDO</a> . Longer term, we can set up an automated service to fetch this regularly from the Internet, although initially we will use JTRIG infrastructure. Some directly requested by CDO via [REDACTED]. |
| <b>known good lists</b>                           | eg, Clean MX (support.clean-mx.de), and perhaps Google's Safe Browsing API could be used (see <a href="#">blog.enjy?</a> )                              | several times a day | none      | small (GB)      | Directly requested by CDO via [REDACTED]   |
| <b>known ORB servers</b>                          | from sources eg, GhostNet   | daily               | none      | very small (MB) | idea from CDO  |

### [edit] Miscellaneous

| Knowledge required  | Available from                                  | Update frequency | Filtering | Volumetrics | Comments  |
|---|---|------------------|-----------|-------------|---|
| <b>UK address to protect</b>  | need to find out how we get them at the moment. | weekly?          | none      | small (GB)  | [REDACTED] apparently got complete list of .gov.uk domains via JANET in June 2011. [REDACTED] trawled <a href="#">KED</a> (and therefore probably Akamai whois data) to find some List X network info.                    |
| <b>USER_AGENT strings, sources, and expected frequency</b>                  | ?   | weekly?          | none      | small (GB)  | see User Agent prototype by [REDACTED]. Of wider interest.  |
| <b>Malware development and hacking techniques being discussed in forums</b> | requires covert monitoring of forums            | weekly?          | ?         | ?           | CKX currently working with E-crime to identify and evaluate forums of potential interest. This project may extend to active monitoring of and reporting on discussions in selected forums. CKX Ops Manager is [REDACTED]. |

POC: [REDACTED]   
 POC: [REDACTED]   
 POC: [REDACTED]

Retrieved from "[REDACTED]"  
 Categories: [Cyber Defence](#) | [Open Source Information](#)

Views

- [Page](#)

- [Discussion](#)
- [Edit](#)
- [History](#)
- [Delete](#)
- [Move](#)
- [Watch](#)
- [Additional Statistics](#)

Personal tools

- [\[REDACTED\]](#)
- [My talk](#)
- [My preferences](#)
- [My watchlist](#)
- [My contributions](#)

Navigation

- [Main Page](#)
- [Help Pages](#)
- [Wikipedia Mirror](#)
- [Ask Me About...](#)
- [Random page](#)
- [Recent changes](#)
- [Report a Problem](#)
- [Contacts](#)
- [GCWeb](#)

Search

Tooltbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)

- This page was last modified on 25 June 2012, at 09:42.
- This page has been accessed 640 times.
- All material is UK. [http://www.gchq.org/organisation/ck/opensource/policy\\_strategy/copyright/](http://www.gchq.org/organisation/ck/opensource/policy_strategy/copyright/) Crown Copyright © 2008 or is held under license from third parties. This information is exempt under the Freedom of Information Act 2000 (FOIA) and may be exempt under other UK information legislation. Refer any FOIA queries to GCHQ on 01242 221491 x30306 or [infoleg@gchq.gsi.gov.uk](mailto:infoleg@gchq.gsi.gov.uk)
- [Privacy policy](#)
- [About GCWiki](#)
- [Disclaimers](#)

TOP SECRET STRAP1 COMINT

The maximum classification allowed on GCWiki is **TOP SECRET STRAP1 COMINT**. Click to [report inappropriate content](#)