

Göran Pestana

Incident handler and developer

CERT  SE

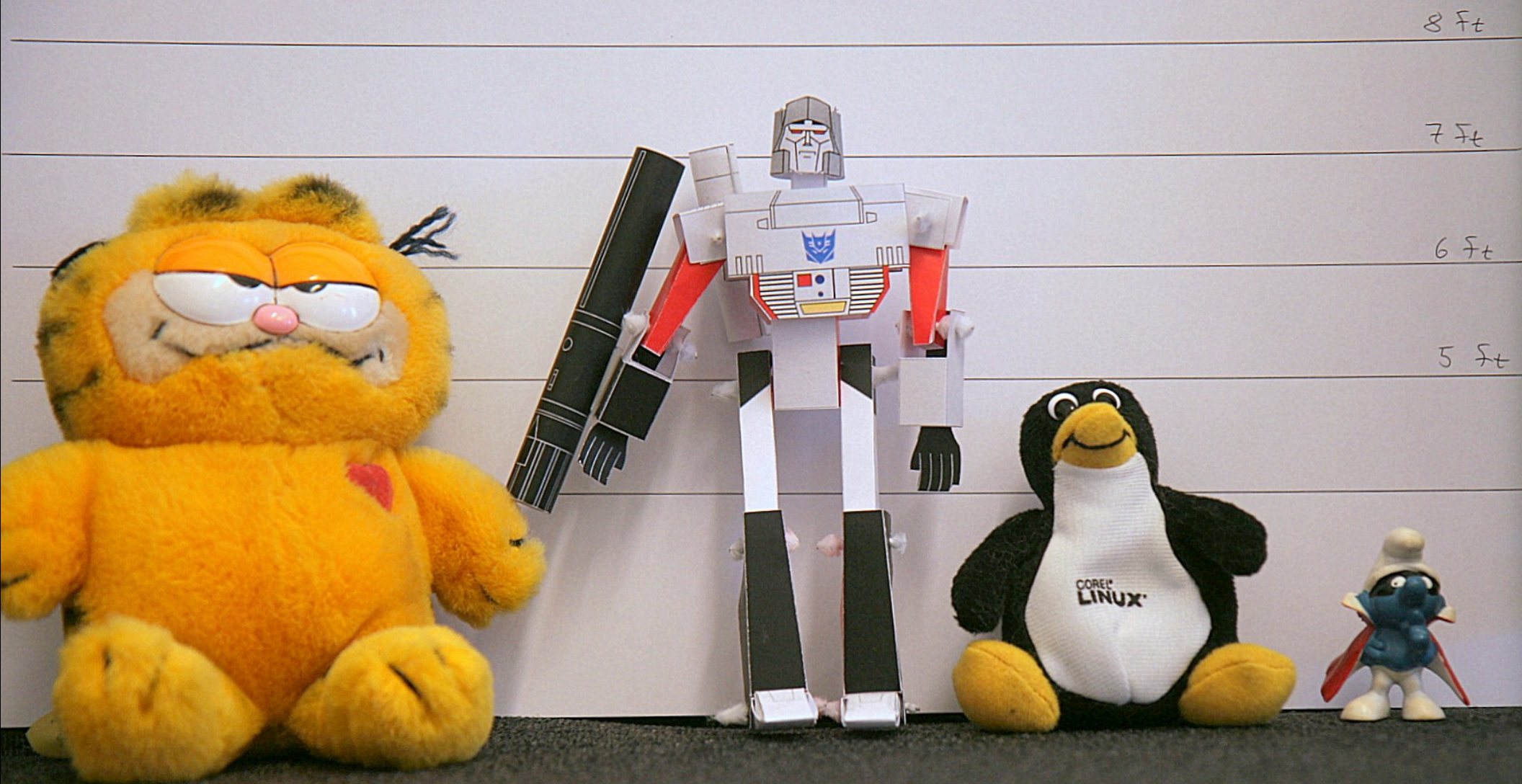
# Megatron

Automated Abuse Handling

by

CERT  SE

# Who is Megatron?



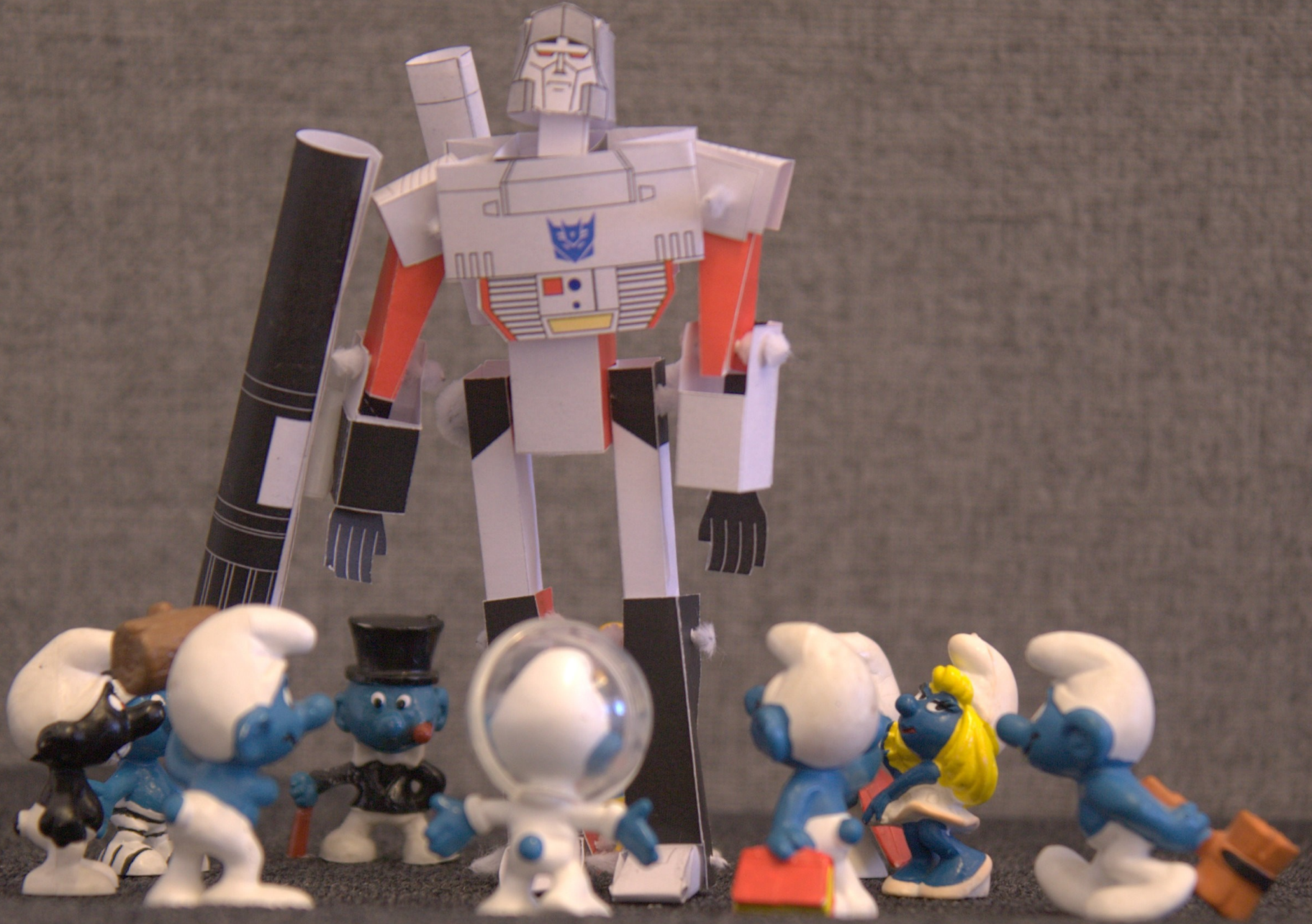
# Who is Megatron?



- A system that collects and processes information about bad hosts on the Internet
  - Input are logfiles from many different sources
  - Checks if the IPs exists in the organization DB
- Developed by CERT-SE
  - Tor Johnson & Göran Pestana
- Implemented in **Java** and uses a **MySQL** DB
- In production since the end of **2009**



# What can Megatron do?



# What can Megatron do?



- Day-to-day abuse handling
- Incident handling
  - Running list of "bad" machines
    - Any matches in the organization database?
    - Send abuse
  - File **conversion**
    - Filtering
    - Add CC, ASN, geolocation, and hostname
    - Other formats
- **Data mining for statistics**

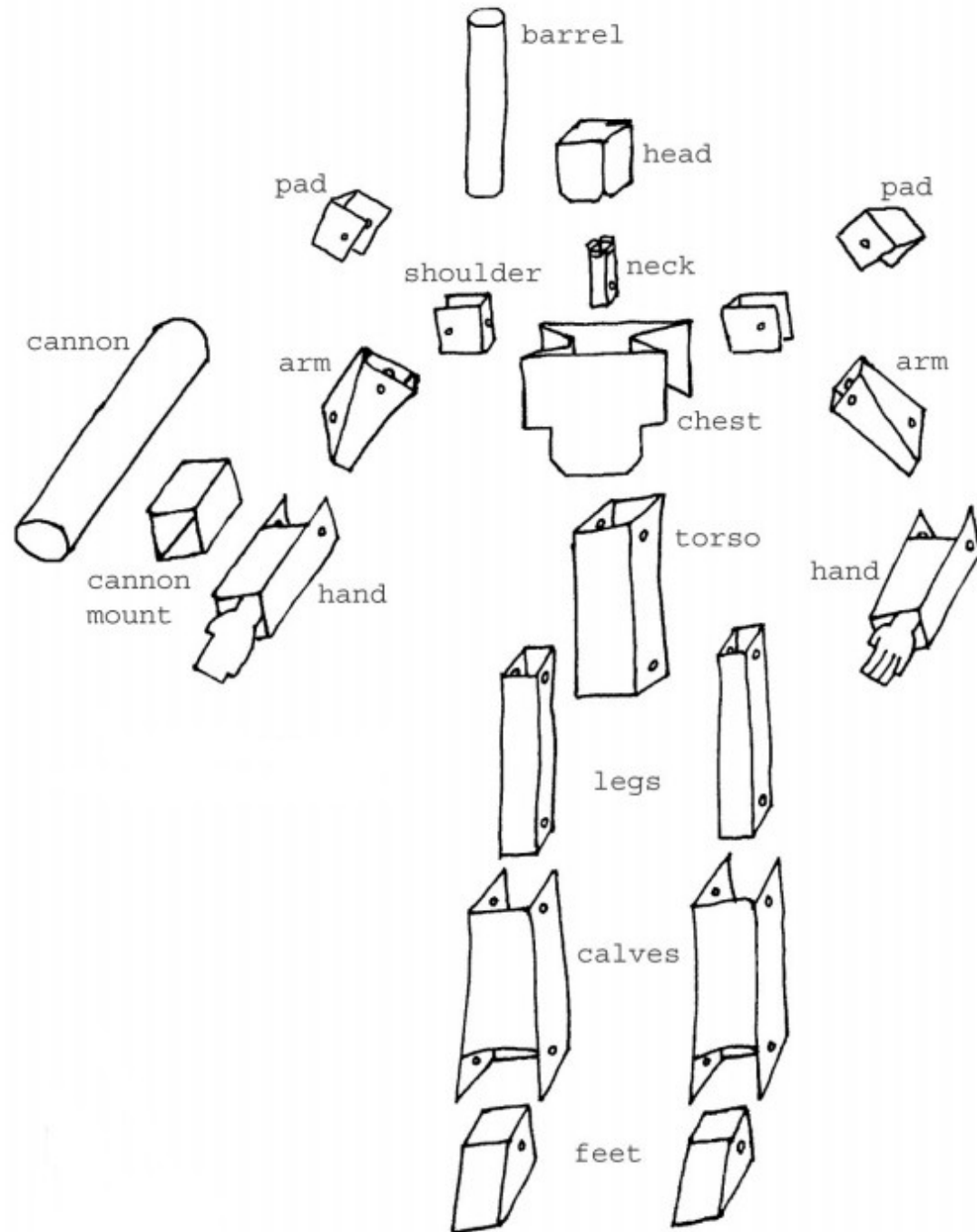
# What can Megatron do?



## Typical statistics for a week @ CERT-SE:

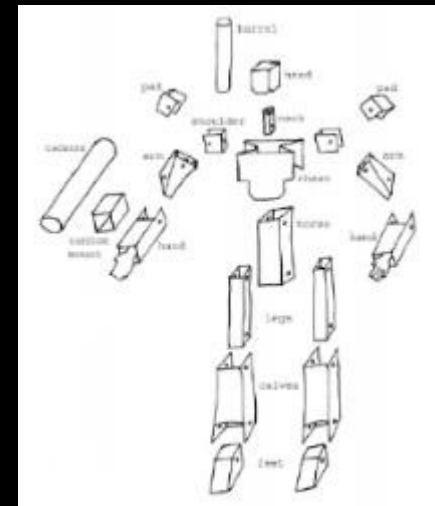
- >10 million log lines processed
- >50 000 log records saved in db
- ~50 abuse emails sent

# Design Goals





# Design Goals

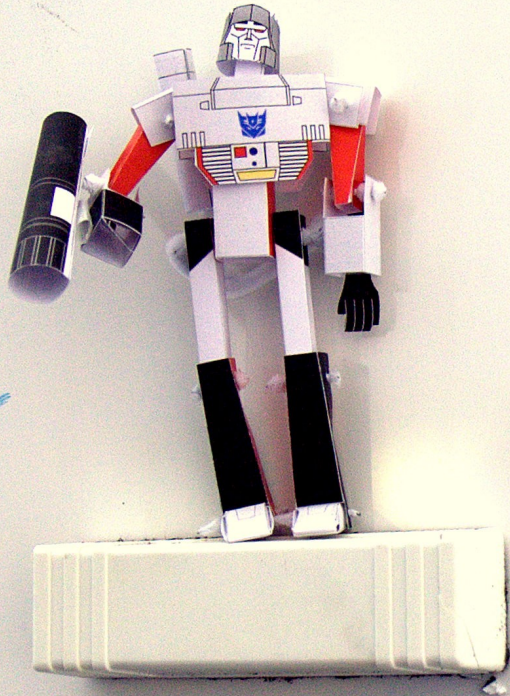


- Should work out of the box
- More than just send abuse mail
- Flexible without coding
- Easy to extend with new code
- Handle large volumes of data
- Fast, only **local** lookups
- Easy to use



# Input Sources

mail  
rsync  
http  
https



RBL  
Shadowserver  
Suhet  
Zeus  
PhishTank



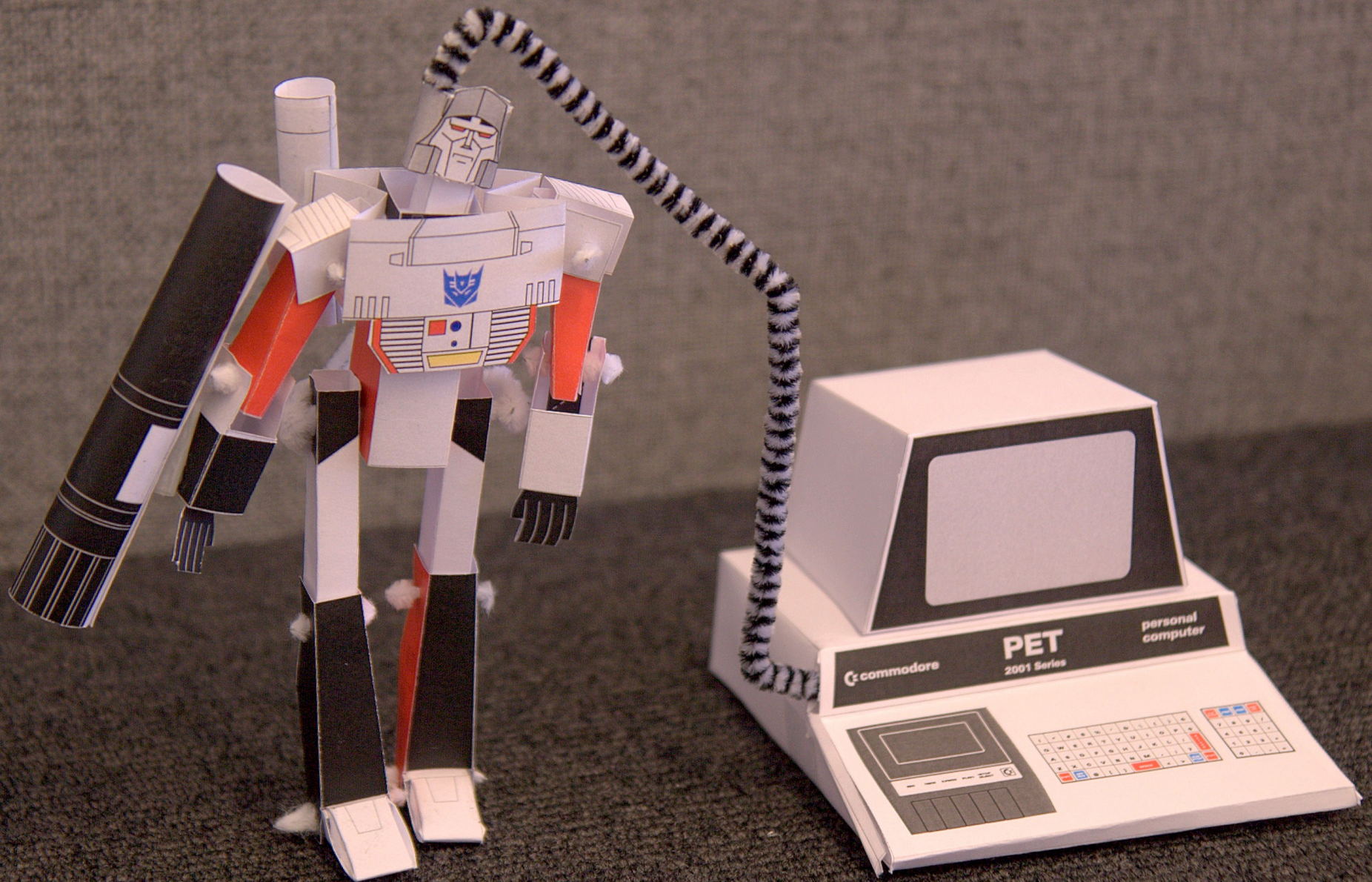
# Input Sources



- **Current:** Shadow Server, Spamhaus, SORBS, Zeus Tracker, DroneBL, Phish Tank, Malware Patrol, Clean MX, Blade Defender, malc0de.com, Zone-h, Turk-h, xssed.com, vs-db.info, SpamCannibal, StopForumSpam, openbl.org, AutoShun, CERT-SE Honey Net, etc.
- ~60 configurations exist so far



# Automation



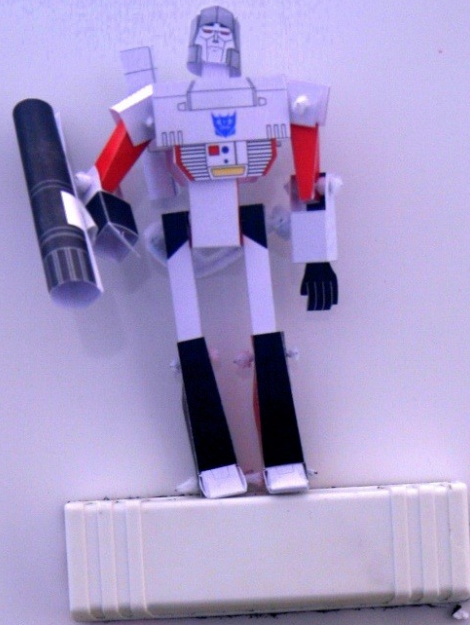


# Automation



- Handler on duty runs megatron and have to approve every mail job
- Helper scripts
  - Download script: `http/https`, `rsync`
  - mailbox-parser: Extracts body or attachments
- "slurp"-directory: Picks config from the filename
- Notification email: Sent if high priority entries are discovered
- RT-integration
  - A ticket-id is created automagic using RT CLI

# Process Steps



Fetch

Matching

Search

Parse

Persist

Export

Filter

Abuse  
Mail

Decoration

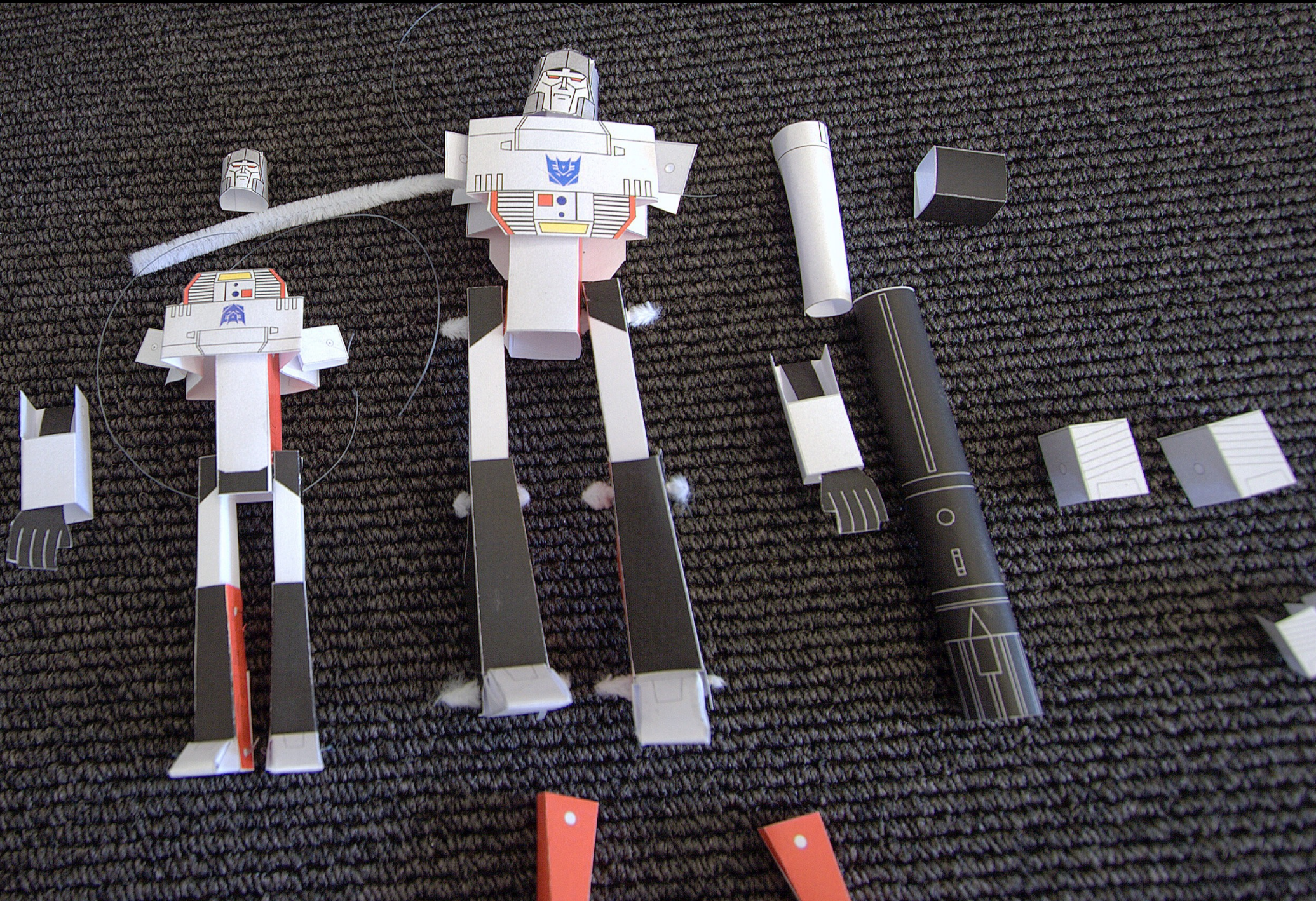
# Process Steps

- **Fetch files:** http/https, rsync, or mail
- **Reformat:** e.g. XML to CSV file
- **Parse:** Tokenize log row
- **Filter:** Before and after every step
- **Data decoration:** Add CC, ASN, and hostname
- **Match:** Bind a log row to an organization
- **Persist:** Store log rows and the parsed data
- **Email:** Send abuse mail
- **File export:** Export data from db using template





# Configuration





# Configuration



- One configuration per input type
- Compact: Global config-properties are inherited
- Parsing using **regexp** and binds result to variables
- A regexp is defined for every variable:

```
parser.item.countryCode=\w{0,2}
```

```
^\w+,"$logTimestamp","$ipAddress",".*?",  
"$additionalItem_infection","$countryCode","$url"
```

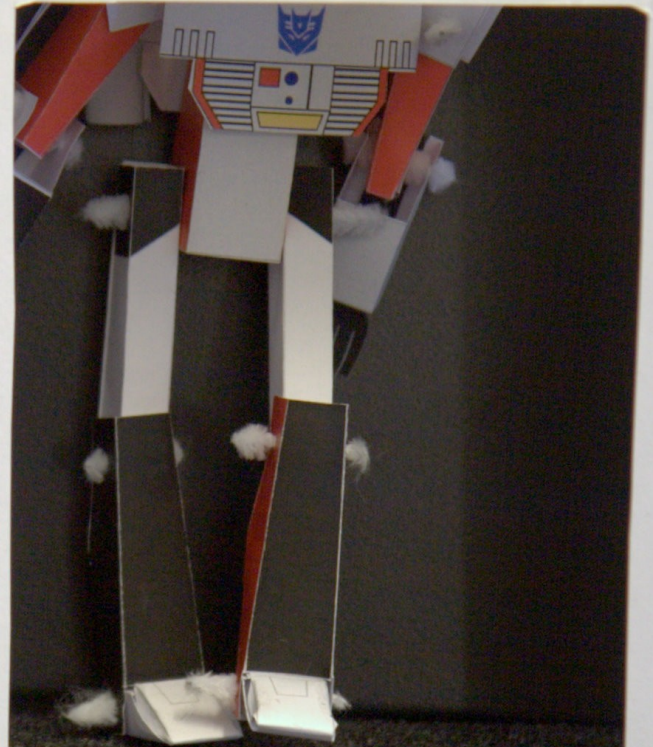
- A variable corresponds to a field in the DB
- Variables are used in mail and export templates

# Filters

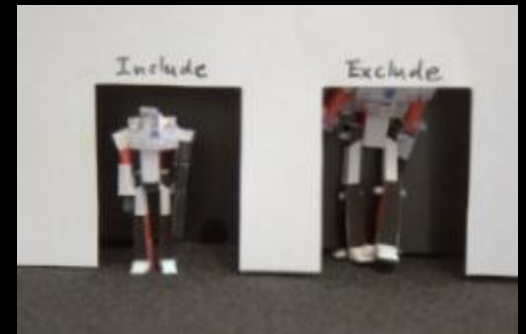
Include



Exclude



# Filters



- Many injection points in the flow
- **Line Filter** (before parsing)
  - Regexp
  - Line number
- **Log Entry Filter** (after parsing)
  - Regexp for arbitrary attribute
  - Country code
  - ASN
  - Priority
- Two modes: **Inclusive** or **Exclusive**



# Data Decoration





# Data Decoration



- **Fast:** Only local lookups and DNS queries
- **Types:**
  - **IP** → **ASN** (BGP data or GeoIP)
  - **IP** → **Country Code** (GeoIP)
  - **IP** → **City, latitude, longitude** (GeoIP)
  - **IP** → **Host Name** (DNS)
  - **Host Name** → **IP** (DNS)
  - Other: **URL** → **hostname**, **hostname** → **CC**
- Dumps BGP table from router and imports it to the Megatron database once a week

# Matching



# Matching



Binding a log record to an organization

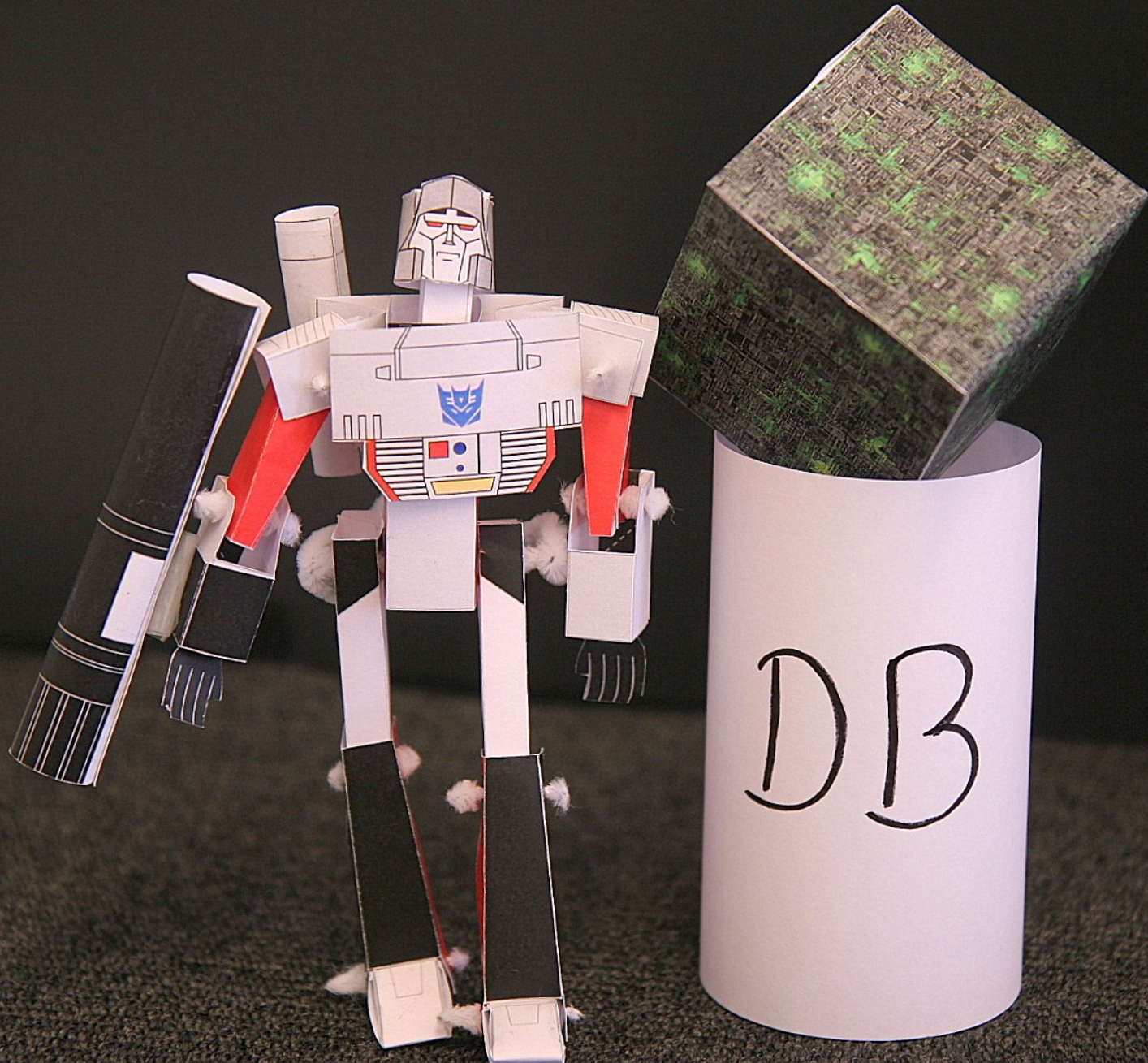
- Governmental agencies
- Municipalities
- Health care organizations
- Critical infrastructure companies
- Etc.

Matching order:

- IP
- Domain name
- ASN

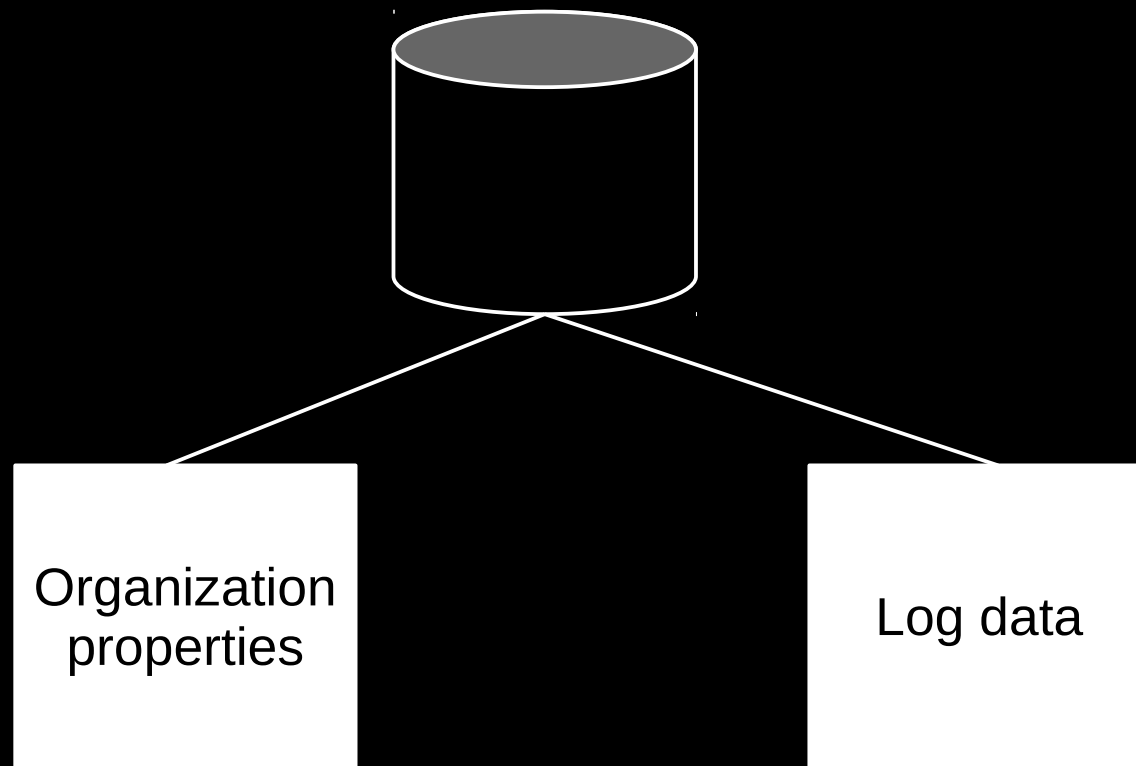


# Database Model





# Database Model

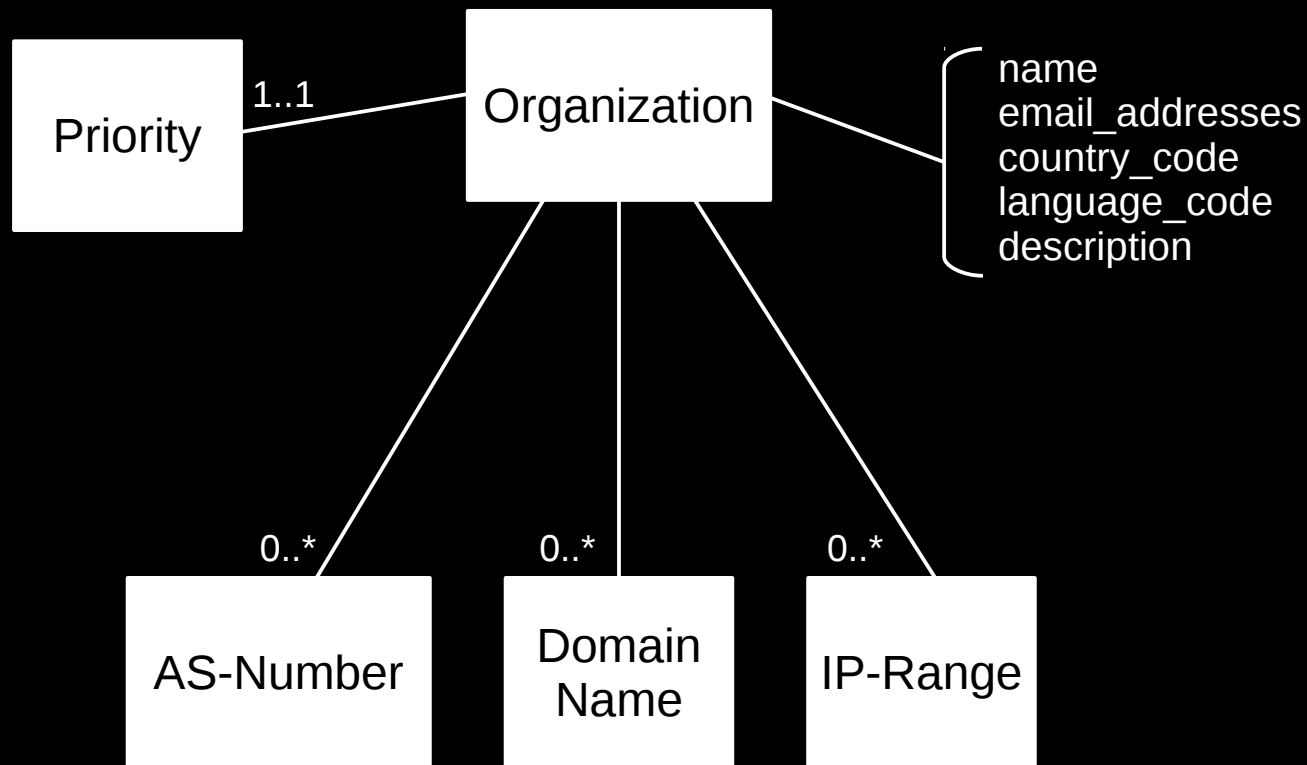


# Database Model

## Organization properties



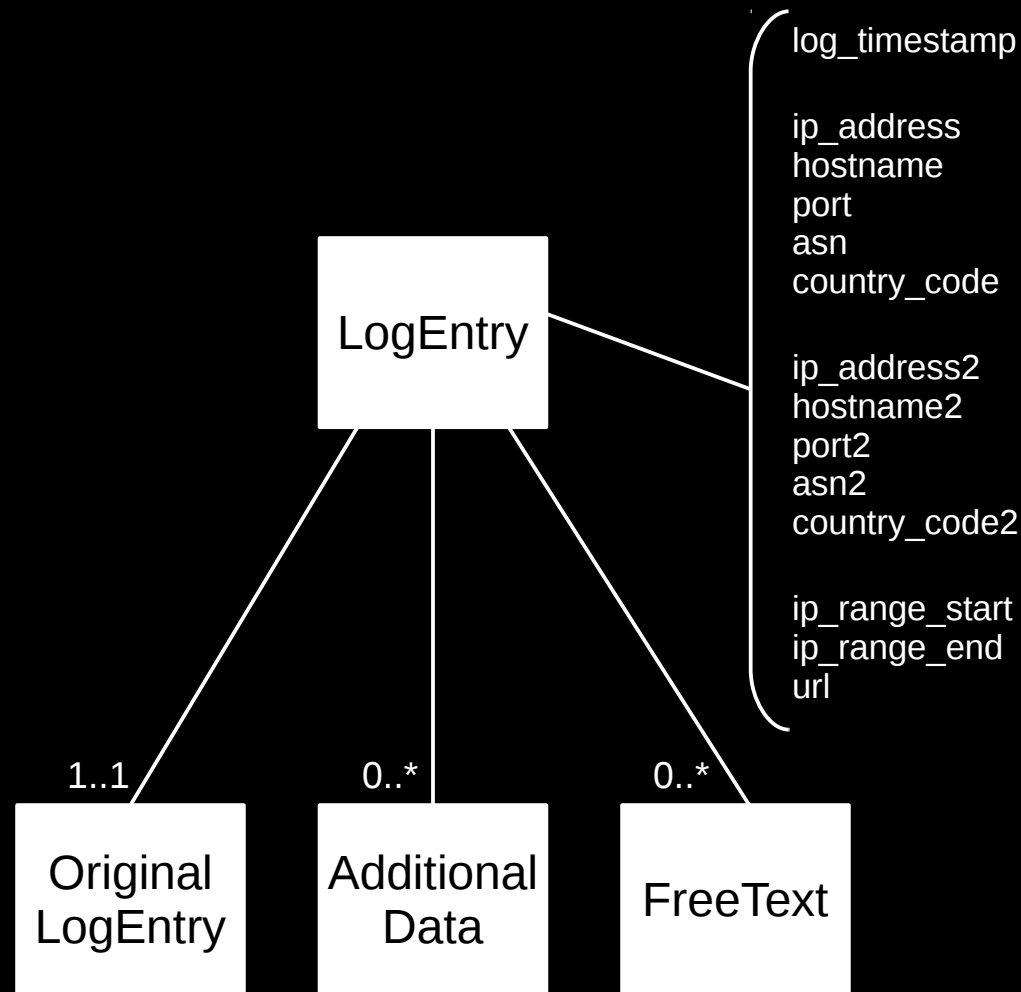
~700 organizations right now



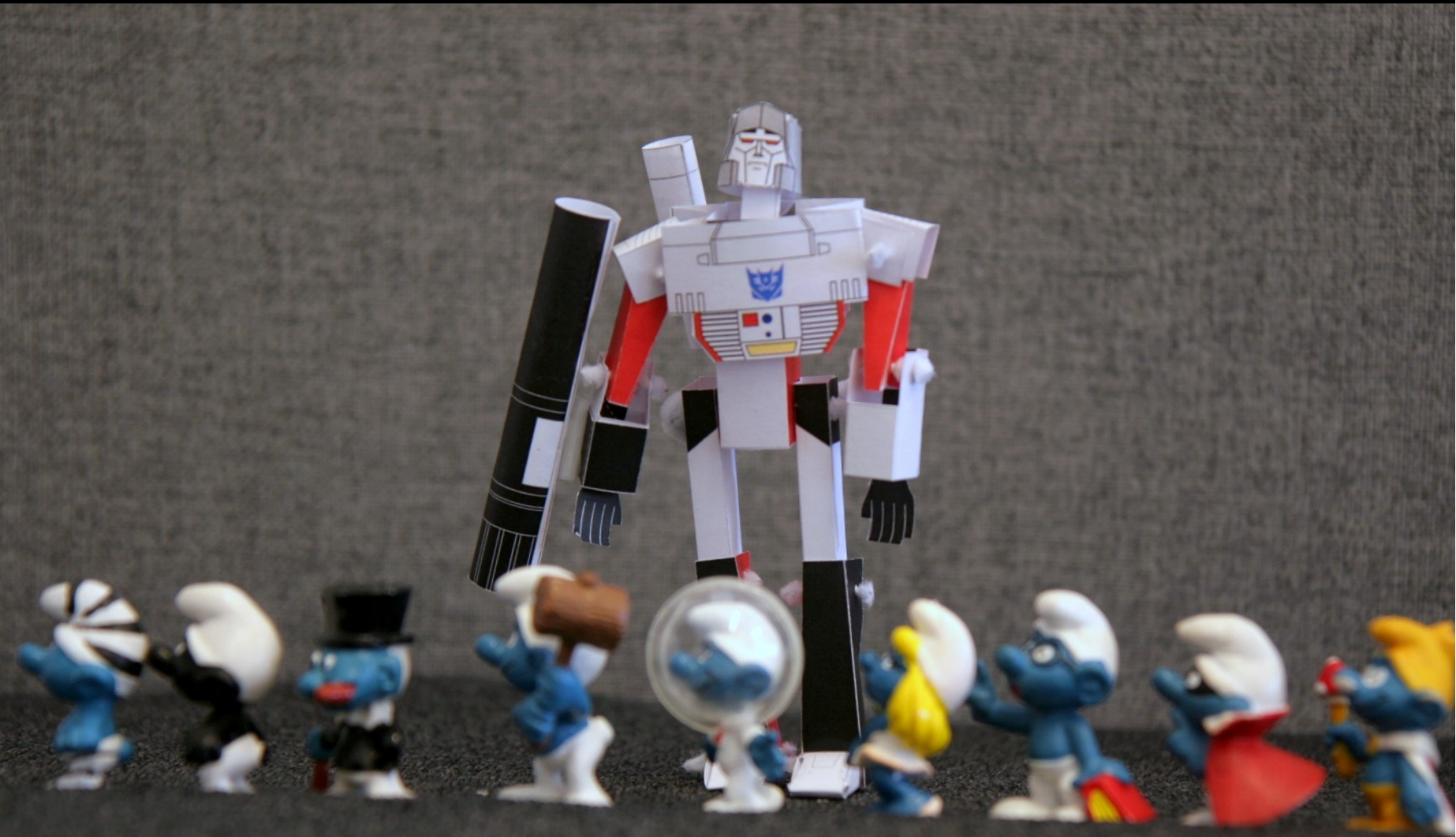


# Database Model

## Log data



# Features





# More Features



- XML-support: Flattens XML files
- Line splitters and merges
- File processor: diffing files
- Checks log-file hash to avoid duplicates
- Support for localized mail templates
- Quarantine IPs to avoid excessive emails
- Flexible configuration of time stamp format, e.g.
  - ISO: yyyy-MM-dd HH:mm:ss z
  - Syslog: MMM dd HH:mm:ss

# Even More Features



- RSS feeds (jobs, statistics)
- Warning for old time stamps
- XML and JSON **data export** for charts



# Where is Megatron?



## Open Source:

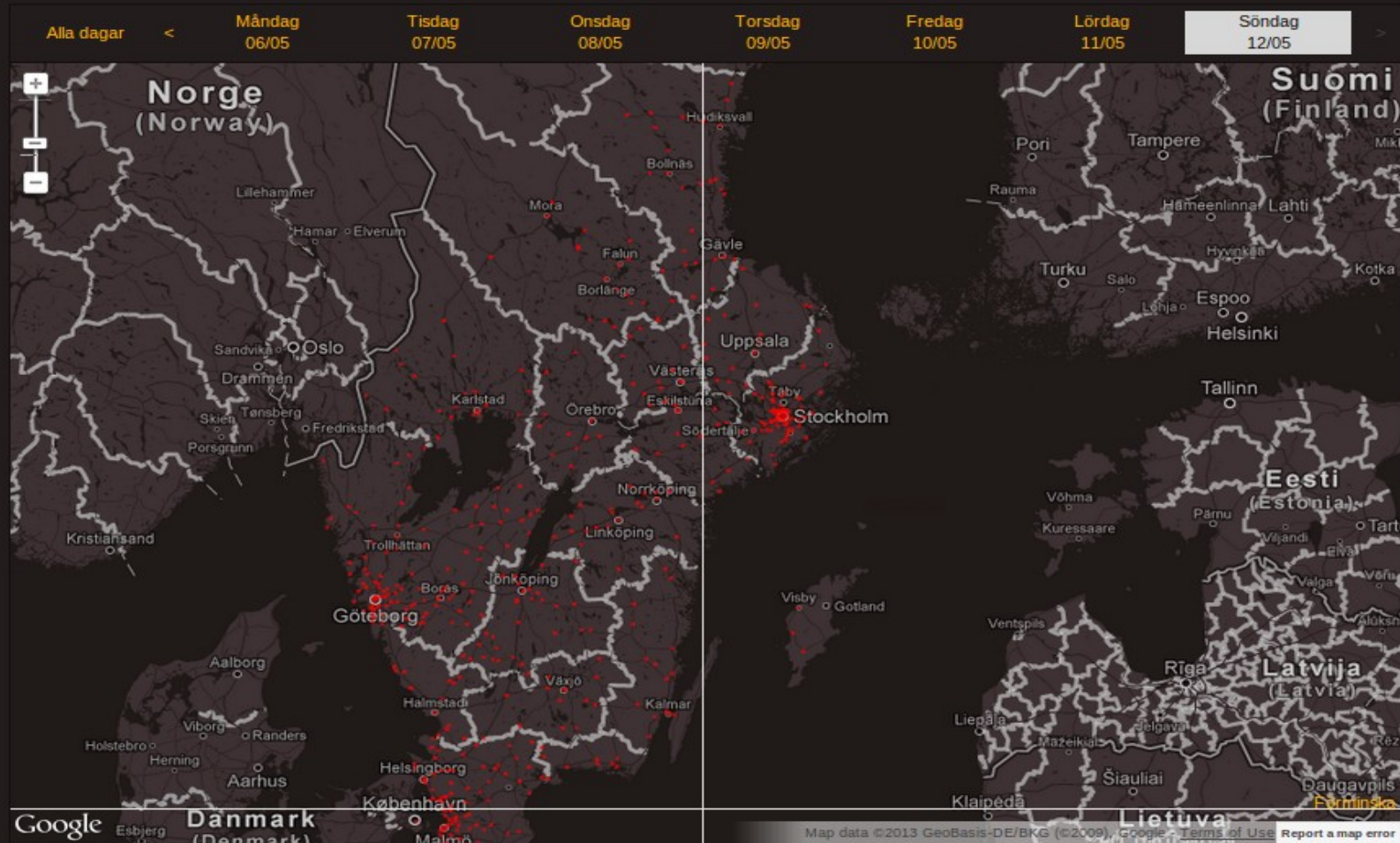
<https://download.cert.se/megatron/>

## Users:

- NCSC-NL
- CERT-HU
- NorCERT
- CERT-CW (Curaçao)
- One major Swedish ISP

# Megamap

## Infekterade datorer i Sverige



### Städer

Stad	IP-adresser	Träffar
Stockholm	7334	31543
Göteborg	2598	10058
Malmö	1810	9088

### Organisationstyper

Organisationstyp	IP-adresser	Träffar
ISP	20026	107958
Webbhotell	13067	61535
ISP 2	3093	6885



# Megamap

- **JavaScript** application
- Static **JSON** files generated by **Megatron**
- Two versions:
  - Public: **www.cert.se**
  - Intranet: IP-addresses unmasked + hostnames
- Benefits:
  - "Eye candy" (upper management **loves** it)
  - Rasing public **awareness**
- It is Open Source:  
Download URL can be provided on request

# Questions?

A paper robot, resembling Optimus Prime, is standing on a small, decorated base. It is holding a large, rolled-up scroll in its right hand. The robot is made of white, red, and black paper. The background is a fiery, orange and yellow pattern.

**Thank you!**

CERT SE