

# OVERVIEW

---

To create an inventory of Internet-accessible assets owned by a single organization, Bit Discovery generates routine snapshots of the entire Internet and uses its patent-pending technology to map relationships between those assets. This report represents our findings.

## 01

### ABOUT

Bit Discovery  
How this was made  
Jeremiah Grossman  
Robert Hansen

## 02

### INTERFACE

Inventory  
Dashboards  
Suggestions  
API

## 03

### TRENDS

Total Assets  
Total Domain Names  
Total Subdomains  
Notable Trends

## 04

### LEXICON

Industry-standard  
terms associated to  
asset inventory and  
related  
technologies.

# ABOUT

## BIT DISCOVERY

Bit Discovery is a cloud-based solution that discovers, inventories, and monitors your internet-accessible assets. Beyond domains and subdomains, Bit Discovery also inventories all technologies it encounters via technology fingerprinting. The world's top companies and agencies trust Bit Discovery to discover just about anything about their internet-facing assets.



Jeremiah Grossman's career spans a literal lifetime of approximately 20 years in information security, during which he has become one of the industry's biggest names. In that time, he has been publicly thanked by Microsoft, Mozilla, Google, Facebook, a variety of Fortune 500 companies, and many others for privately informing them of weaknesses in their systems and relieving many of their information security woes. This experience and access have given Jeremiah unique insight into some of the day to day struggles of today's dynamic online business world. Internet asset management was a common challenge with no good solution in sight. The demand for Bit Discovery has been growing as fast as the Internet, and Jeremiah heeded the call. Jeremiah Grossman has also founded WhiteHat Security, is a Brazilian Jiu-Jitsu black belt, published author, influential blogger, and off-road race car driver. <http://www.jeremiahgrossman.com>



Robert "RSnake" Hansen is the data-whisperer who brings all of Bit Discovery's "bits" together. Another information security lifer, Mr. Hansen started his career at eBay, where he was responsible for authentication as well as most anti-fraud systems and anti-phishing technologies. Since then, Robert has worked in nearly every bit of web application security for many of the world's largest organizations, including over 2,100 banks, credit card processors, flight control systems, SCADA (water and power) control systems, and other security companies. It is safe to say that nearly no one knows the Internet as well as he. In his spare time, Robert is a frequent keynote speaker and is on the speaker selection committee for Black Hat and Hack in the Box.

# ASSET

as·set | 'aset |

noun

a domain name, subdomain, or IP addresses and/or combination thereof of a device connected to the Internet or internal network.

- (an **asset**) may include, but not limited to, web servers, name servers, IoT devices, or network printers.

# THE DATA

## BEHIND THIS REPORT

# 4.5 Billion

## ASSETS

The most comprehensive view of the Internet. Well beyond just DNS entries and IP-addresses, we capture relationships between Internet-accessible assets.

**200+**  
INTERNET  
SNAPSHOTS

Each internet snapshot contains details about every known DNS entry and IP-address (a.k.a. Asset) we're aware of across the entire public Internet. Snapshots are timestamped and have been generated many times per year for the last 5 years, enabling Bit Discovery to closely track changes about all Internet-accessible assets.

**515**  
DATA  
SOURCES

Each Internet snapshot is enriched with third-party data including WHOIS databases, registered top-level domains, ASN information, listening ports, service banners, technology stack details, SSL/TLS certificate info, RBL lists, email addresses, password dumps, vulnerability data, etc.

**115**  
DATA  
COLUMNS

Data collected about each and every asset in every Internet snapshot has potentially up to 115 unique columns of data. A data column, often referred to as asset meta-data, may include the web server distribution and major/minor version, CMS type, TLS certificate expiration date, geo-IP physical location, cloud or CDN provider, etc.

**150**  
YEARS OF  
CPU TIME

We routinely process hundreds of terabytes of data, which results in every imaginable compute problem such as disk utilization, bandwidth, CPU, memory, and I/O processes. Processing Internet snapshots require a highly customized and purpose-built hardware and software platform split between public and private cloud.

# CAVEATS

## Data Collection

- Our Internet scanners sometimes use ANY type lookups and not all service providers support ANY type DNS lookups (i.e. Cloudflare)
- Round Robin DNS sometimes finds a lot of assets, sometimes a little, and changes frequently.
- DNS servers and resolvers sometimes experience outages.
- DNS responses may exceed TTL.
- DNS servers may selectively blocking requests.

## Accuracy of an Organization's Asset Inventory

- Assets with subdomains within the ownership of a third-party domain (e.g. <company>.wpengine.com) may cause issues.
- Assets not listed on certificate transparency and/or doesn't have a public DNS entry (e.g. they'll use internal DNS and a self-signed cert).
- DNS errors falling outside the RFC standard, "example\_site.com" (~1%)
- Wildcard (\*) DNS entries.
- DNS providers respond with erroneous information due to breach.
- WHOIS redaction due to GDPR.