We Go Deeper

General Information & Data Collection Review

Profound Networks, LLC

- Founded in 2005

- Address:
  Profound Networks, LLC
  227 Bellevue Way NE, PMB 900
  Bellevue, WA 98004

- Website: www.profound.net

- Ownership: privately owned and self-funded WA State Limited Liability Company

- Compliance Contacts:

  o Van Zell, COO van@profound.net

  o Gavin Carr, CTO gavin@profound.net

  o Matt Wilmot, Attorney, matt.wilmot@stoel.com

General Description: Profound is a leading technographic data provider that has developed a global comprehensive and consistent process to empirically collect an in-depth census of the Internet up to twice monthly or at a cadence requested by clients. This insight includes more than 300+ public-facing attributes attached to a domain that Profound refers to as Digital Business Intelligence (DBI). DBI represents technographic data on the digital infrastructure of a company's public facing network that includes XaaS providers, web analytics, unique devices, cyber security and network growth signals to name a few attributes. Historical data is available for up to eight (8) years. No data is collected from behind a company's firewall. Each data asset packaged for license is vetted and assured to result from publicly available (but difficult to find) information.

- Link to Profound's Product Reference Guide:
  www.profound.net/files/Profound_Reference_Guide2020.pdf

- Data Sources: The data can be broken into four (4) categories:

  o Public data – Includes data derived from public DNS, public website HTML source files, licensed Whois directory, website content, IP fingerprinting data

- o Private 3rd party data: Certain data attributes sourced from companies including:

    - Dun & Bradstreet

    - Moody's/BvD

    - Experian

    is limited to users, with verifiable license agreements in place with these providers. No 3rd party data is included in Profound deliverables unless specifically requested by our client and, where required, authorized processor's agreements are in place.

    3rd party data may include: D&B DUNS Numbers, BvD ID or Experian BIN numbers, plus firmographic data such as SIC/NAICS codes, number of employees and annual revenue. 3rd party fees may apply at the discretion of each bureau.

  - o Derivative scores & firmographics– Scores are generated by Profound and are based on the data that Profound collects. Profound has an ongoing effort to collect firmographics to supplement our technographic data collection leveraging machine learning.

  - o Additionally, Profound periodically licenses 3rd party sources of firmographics for inclusion in our data offerings.

- Data Harvesting: Profound's processes include web indexing with all of our data collection tools being source transparent. The majority of our data is collected from HTML source which is not visible to the casual website visitor. We have internally set limits on bandwidth to avoid adversely impacting the operation of website where we are collecting data. We do not obscure/ghost our identity when scraping. There is no consumer PII or HIPAA data collected. Profound has a formal opinion on the compilation of Whois data and continuous counsel regarding all of its compilation practices. As an example, the consensus regarding Profound's compilation of DNS data (a very significant part of our data collection assets) concludes that there is no one jurisdiction that controls the DNS and that there are no restrictions collecting DNS data. Profound honors all cease and desist letters. There have been less than a handful of requests from companies to remove their signal from our data profiling which we have respected.

  Profound does not use APIs from target websites. We collect a vast array of data from 310 million unique websites and do not focus more resources on any particular site over another.

  Profound complies with /robots.txt which is how websites manage their automated indexing apps. such as what Profound deploys.

It is impractical for Google, Bing, Yahoo, Baidu, and every other search engine to read and interpret the terms and conditions for every single website they index, so the **robots.txt** page is what the industry has come up with to handle web indexing engines. If a site owner wants nothing from their site indexed, and they want to remain practically invisible on search engines, all they need to do is have a robots.txt file with the following rule:

user-agent: *

Disallow: *

Most websites that are concerned with data mining and website indexing maintain a file on their webserver called "robots.txt" this file provides rules to website index engines (bots) as to which pages, they are free to index and which pages they are not allowed to index. Every company with a website can fine tune its preferences for which pages bots can index and which ones they do not want indexed by modifying this file.

As a pertinent example, if you go to www.dnb.com/robots.txt you will find

the following (page current as of December 30th, 2021):

```
User-agent: *
Disallow: /benefits/
Disallow: /development/
Disallow: /c/
Disallow: /references/
Disallow: /personalization/
Disallow: /investor-relations/
Disallow: /errorhandler/
Disallow: /content/dnb-mod/
Disallow: /content/dnb-us/
Disallow: /content/dnb/
Disallow: /content/dnb-uk/
Disallow: /content/dnb-ca/
Disallow: /content/dnb-ie/
Disallow: /content/dnb-collections/
Disallow: /content/connectors/content/
Disallow: /*/jcr:content/
Disallow: /jcr:content/
Disallow: /online-terms/
Disallow: /trade-partner/
Disallow: /content/dam/english/private/
Disallow: /site-search.html
Disallow: /site-search-new.html

Sitemap: https://www.dnb.com/sitemapindex.xml

User-agent: ScoutJet
Disallow: /

User-agent: Twitterbot
Disallow:
```

DNB.com specifically calls out ScoutJet bot to **not** index ANYTHING at all from their site. DNB.com then Disallows the sections that they do not want any bot to visit at all.

All other bots are free to visit the pages that have not been specifically Disallowed and to index content from those pages.

Profound's bot would find and visit…

http://www.dnb.com/company/contact-us.html

…Which includes the HQ address for D&B. This page is not a page Disallowed by robots.txt,

Therefore, D&B has essentially granted us (and everyone else) permission to index this page.

Included in the HTML source of this contact page, and most of the other pages on DNB.com are data attributes that we collect as part of our DBI suite. For example, if you look at the HTML source, we see:

www.google-analytics.com

Which tells us that DNB.com uses Google Analytics.

We see: munchkin.marketo.net

Which tells us that DNB.com uses Marketo

And many other technographics that are all perfectly legible by looking at the HTML source which all browsers must see in order to render each page. Nothing proprietary here, no trade secrets, this is all PUBLIC data not Disallowed to be indexed by robots.txt. Profound programmatically parses this metadata and incorporates it into our DBI Suite.

All legitimate web indexing engines (including Profound's) respects the rules outlined in robots.txt.

> Profound will look at tens of millions of different websites, indexes the organizational level location information, publicly available technographics and moves on. We have received less than 5 inquiries about our indexing over the past three years after literally hundreds of millions of sites indexed.

> No Personally Identifiable information (PII) is collected. And we are meticulous when it comes to respecting international law, minimizing our footprint, respecting network admins and the resources they look after.