

## SourceScrub Differences

"SourceScrub has the most comprehensive universe of bootstrapped companies' data, more so than any of their 'competitors' and has allowed us to scale our sourcing efforts exponentially. The bootstrapped database is really the selling point. "

—Veljko Bogdanovic, VP,  
Totality Ventures

"SourceScrub has the most complete data set from all the vendors we evaluated"

—Managing Director,  
Mid-Market PE Firm

# Covering the Private Company Landscape

Understanding the universe of investment-ready, founder-owned companies is extremely complex. Private company data is spread across thousands and thousands of highly disparate, heterogeneous and constantly changing sources. And dealmakers need more than just company names. To find, qualify, prioritize and engage targets, they need extensive profile data, accurate contact information and list sources.

Validating, organizing, and continuously updating millions of data points requires a sophisticated combination of technology, people and processes.

## The SourceScrub Data solution

SourceScrub's data engine and operations were built to solve the unique challenges of private company intelligence. They combine cutting edge machine learning with advanced web technologies and an 800-person data operations team. The result is the most accurate, complete and fresh private company data set available.

### The best of people and technology

Effective machine-learning and artificial intelligence (ML and AI) requires human-assured interpretation to accurately scale data aggregation and classification. SourceScrub's **800-person data operations team** not only supervises the machine learning model to ensure data fidelity, but also interprets and hand-writes company descriptions so they depict the business more accurately than simply scraping website copy..

### Reliably fresh, accurate, and complete data

SourceScrub company profiles include a minimum of **7 characteristics** to guarantee companies are real, and not just websites. This mosaic of data also allows users to not only identify targets, but also understand their relative stature and growth trajectory. Because profile data is dynamic, we refresh data points as often as every 30 days, providing trended data from inception.

### Exhaustive coverage of the private company universe

SourceScrub provides the highest volume and variety of sources and lists available, cross-referenced with company data. **115,000+ sources** such as industry conference lists, best-of lists, award publications, top lists, third-party market maps, buyer's guides, professional association lists, and transaction portfolios give the user a much more complete and objective view of markets.

# How it works

## Overlapping Methods Ensure Comprehensive Coverage

The innovative foundation of SourceScrub data lies in the way we collect and index hundreds of millions of private company data points.

### 1. Contributory Curation:

Our vast repository of data channels is curated from direct and indirect requests from thousands of active SourceScrub users who submit lists and companies for our research operations team to vet, validate, and curate.

### 2. Unique Discovery:

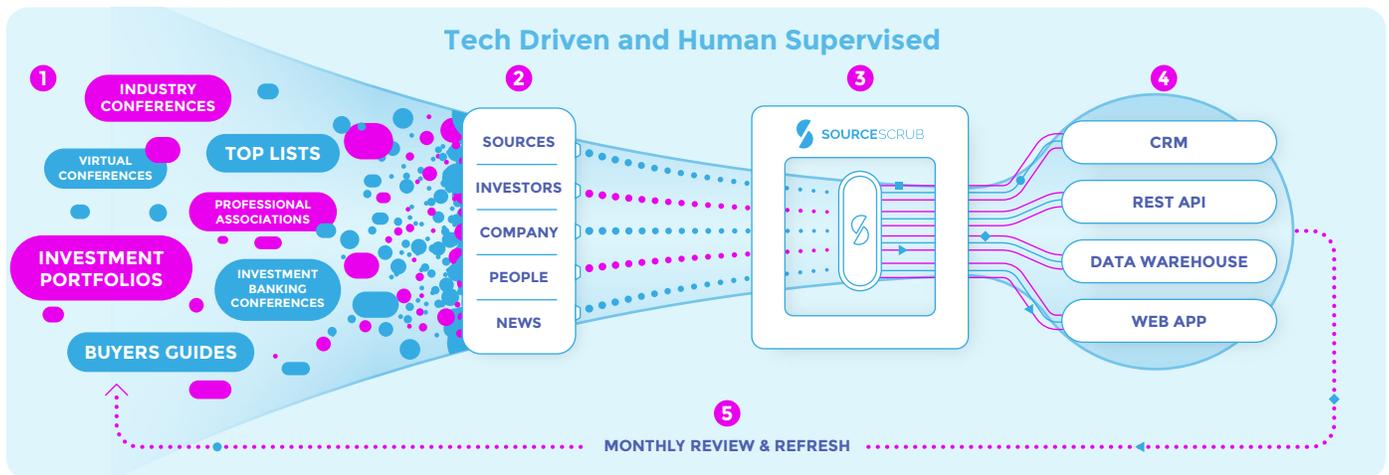
Contained within lists are the difficult-to-find private companies. Thematic connectivity across companies and sources drives additional database population within sub-sectors.

### 3. Complete Company Profiles:

SourceScrub requires a minimum of 7 characteristics before a company profile is published.

## 5-Step Process Delivers Unparalleled Accuracy

Our 5-step data process ensures completeness, accuracy, and freshness.



### 1. Collect:

Crawlers and researchers work methodically across 115,000+ sources to find and ingest data.

### 2. Structure:

Normalizing and structuring data is critical to get the data model right, and ensures users can create the right connections across companies and across verticals. Our dimensional data model creates linkages across companies, sources, investors and people.

### 3. Cleanse:

Data analysts QA and enhance data in a way technology alone cannot. The data operations team assists with labeling and interpretation, which paves the way for machine learning and training to scale this process.

### 4. Deliver:

SourceScrub's engineering and product team package data so it is accessible in a range of formats, including: web platform, web extension, CRM integration, direct data feeds into data warehouse, or REST API calls into proprietary applications.

### 5. Refresh:

Stale data can be misleading. SourceScrub's data set is reviewed monthly and data points are refreshed and QA'ed so SourceScrub data is consistently fresh and accurate.