

The External Data Market Map, 2026

Four ways companies get data they don't own, and what each one costs them.

Forage AI · 145 companies · Edition: October 2026

1 Buy it ready

YOU GET

Data that already exists, fast

AI SHIFT Vendors now plug their data straight into AI assistants through MCP servers

Company & people data 9

Apollo.io

Cognism

Coresignal

Crunchbase

Dun & Bradstreet

Lusha

People Data Labs

PitchBook

ZoomInfo

Financial & alternative data 8

Consumer Edge

M Science

Nasdaq Data Link

Placer.ai

Preqin

S&P Global Market Intelligence

Similarweb

YipitData

Healthcare data 5

Definitive Healthcare

H1

HealthLink Dimensions

IQVIA

Komodo Health

Real estate & property data 6

ATTOM

CoStar

Cotality

RealPage

Regrid

Reonomy

Retail, ecommerce & pricing data 6

Circana

DataWeave

NIQ

Profitero+

Stackline

Worldpanel by Numerator

Open & ready-made datasets 4

Bright Data Datasets

Common Crawl

Hugging Face Datasets

Kaggle Datasets

Data marketplaces & exchanges 7

AWS Data Exchange

BattleFin

BigQuery sharing

Databricks Marketplace

Datarade

Eagle Alpha

Snowflake Marketplace

2 License it

YOU GET

The legal right to use content its owner controls

AI SHIFT AI turned content licensing from one-off deals into a metered market

Licensing marketplaces & intermediaries 8

Copyright Clearance Center

Created by Humans

Dappier

Human Native

Microsoft Publisher Content Marketplace

ProRata.ai

RSL Collective

TollBit

Paid access at the network edge 5

Akamai

Cloudflare AI Crawl Control

DataDome

Fastly

Skyfire

Licensed training-data brokers 8

Defined.ai

Dow Jones Factiva

Getty Images

Protege

Shutterstock

Troveo

Wikimedia Enterprise

Wirestock

3 Build it yourself

YOU GET

Full control over sources, schema and timing

AI SHIFT AI made the first build cheaper; running it in production still lands on you

Proxies & unblockers 7

Bright Data

Decodo

IPRoyal

Oxylabs

SOAX

Webshare

Zyte API

Agent browsers 7

Anchor Browser

Browser Use

Browserbase

Browserless

Hyperbrowser

Kernel

Steel

Search APIs for AI agents 8

Brave Search API

Exa

Linkup

Parallel

Perplexity Search API

SerpApi

Tavily

You.com API

Crawl & extract APIs 10

Diffbot

Firecrawl

Jina Reader

Nimble

Olostep

ScrapeGraphAI

ScraperAPI

ScrapingBee

Spider

ZenRows

Scraping platforms & no-code tools 7

Apify

Browse AI

Import.io

Kadoa

Octoparse

ParseHub

Thunderbit

Document parsing software 8

ABBYY

Hyperscience

Instabase

LlamaParse

Mindee

Nanonets

Reducto

Unstructured

Open-source frameworks 7

Crawl4AI

Crawlee

Docling

Playwright

Puppeteer

Scrapy

Selenium

4 Have it built

YOU GET

Data built to your spec, kept running by someone else

AI SHIFT Providers use AI inside the pipeline; the accountability for the output stays human

Managed web data services 9

Actowiz Solutions

Bright Data Managed Services

Datahut

Forage AI ♦

Grepsr

PromptCloud

ScrapeHero

X-Byte

Zyte Data

Managed document & data processing 4

DataEntryOutsourced

Hitech BPO

Innodata

SunTec India

Custom AI data creation (human data) 12

Appen

Handshake AI

iMerit

Invisible

Mercor

micro1

Sama

Scale AI

Surge AI

TELUS Digital

Toloka

Turing

Method: live commercial offering and 2026 ownership checked for every company on 7 Oct 2026. Acquired companies listed under the acquirer. Not exhaustive. No paid placement.

♦ marks the author, Forage AI · forage.ai/blog/external-data-market-map