DataCrawler

Rancho's DataCrawler

Rancho's DataCrawler lets users crawl, manage and enrich data and metadata from PubMed, ClinicalTrials.gov, and other commonly used public resources. "Light Crawler" and "Deep Crawler" modes are customizable to your project's needs to provide an efficient yet comprehensive search experience. By automating the data extraction process, DataCrawler can save time, reduce costs, and provide valuable insights that can lead to improved decision-making and competitive advantage.
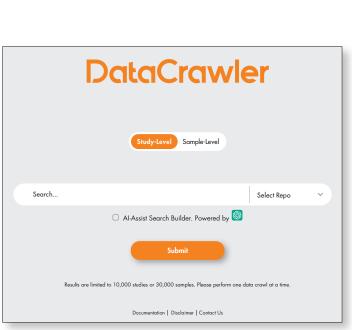
FEATURES:

Easy to use interface eliminates the need for complex scraping scripts and development time. Simply input your search and a selected repository and walk away.

Access to millions of research papers, specimens, and datasets

Annotation Service – automatically annotates keywords in free-text fields from both metadata and sample levels against public ontologies.

Customizable to extract data from various websites; can be used on static and dynamic web pages.

Can extract data in different formats, including JSON, CSV, and Excel, for easy integration into various analytical tools and platforms.

DataCrawler

Study-Level   Sample-Level

Search...   Select Repo ⌄

☐ AI-Assist Search Builder. Powered by 🟢

Submit

Results are limited to 10,000 studies or 30,000 samples. Please perform one data crawl at a time.

Documentation | Disclaimer | Contact Us

AI:

- AI-Assist enhances your search queries by optimizing them into Boolean strings and augmenting them with gene and disease synonyms to ensure you capture all essential data. Additionally, it customizes your search strings for a specified repository.

- AIScore rates each returned study based on relevance to your search on a scale of 1-100, facilitating easy prioritization for reviewing results.
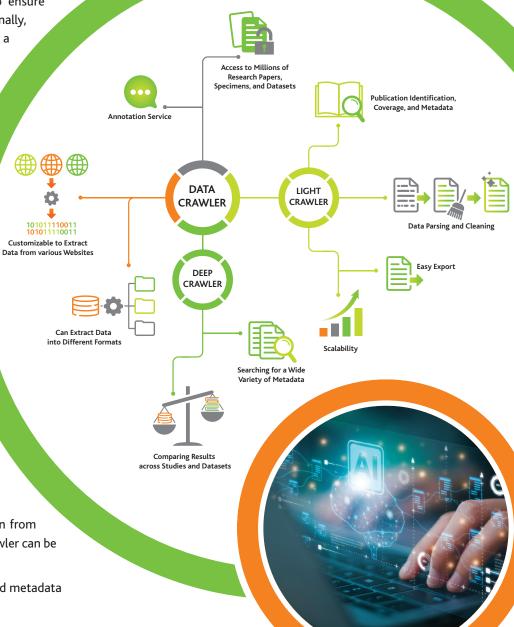
DataCrawler enables cross-referencing with publications (PMID) and other closely related data sources and link corresponding data accession identifiers. This feature enables users to easily discover related resources and avoid pulling redundant data from different data sources. For example, for datasets found in GEO data archive, Crawler will cross-reference to SRA or BioProject if the corresponding data are also archived in these data repositories.

LIGHT CRAWLER

- Lets users extract valuable information from their target publications. The Light Crawler can be used for:

- Publication identification, coverage, and metadata

- Data parsing and cleaning

- Easy export

- Scalability

DEEP CRAWLER

Designed to gather sample-level metadata from publications. The Deep Crawler can be used for:

Annotation Service

Access to Millions of Research Papers, Specimens, and Datasets

Publication Identification, Coverage, and Metadata

101011110011
101011110011
Customizable to Extract Data from various Websites

DATA CRAWLER

LIGHT CRAWLER

Data Parsing and Cleaning

Easy Export

DEEP CRAWLER

Can Extract Data into Different Formats

Scalability

Searching for a Wide Variety of Metadata

Comparing Results across Studies and Datasets

- Searching for a wide variety of metadata, including (but not limited to) sample identification numbers, sample dates, sample processing methods, and experimental conditions

- Comparing results across different studies and datasets