Datasets for drug discovery

Discover comprehensive, trusted, machine-readable data to:

- Inform virtual screening
- Reveal promising drug targets
- Improve biomarker identification and prioritization
- Assess the current treatment landscape for a disease
- Make go/no-go decisions using predictive chemistry.

20M articles

2,900 journals

10 therapeutic areas

18M biological relations

9M substances with bioactivities

56M substances

Full-text journals data

- 24 Major Disciplines covered; including Biochemistry, Genetics, Chemistry, Pharmacology and Toxicology.
- High-impact research from *Cell Press*, *Lancet* and more.
- Therapeutic area data subsets: Cardiovascular, Dermatology, Oncology, Endocrine, Diabetes & Metabolism, and more.

Biological relationship data

- Biological relations from full-text and MEDLINE abstracts.
- Small molecule protein interactions.
- Relations from clinicaltrials.gov.
- MiRNA effects, SNP annotations and more from public datasets.

Chemistry data

- 23M single-step, full reactions.
- 9M substances with associated bioactivities (target and assay information).
- 56M substances (fact availability, patent references).
- Complete dataset access via API.

Numbers as of September 2024, rounded up, and may vary with delivery mechanism.

ELSEVIER

Quality data

- Peer reviewed, highly cited article content.
- Broad biomedical and chemistry coverage: medicine, biochemistry, genetics, pharmacology, toxicology and chemistry.
- Human expertise in manual excerption, data science and the scientific domain ensures data accuracy.
- Continual investment in AI/ML excerption drives quality data.
- Clear provenance for evidence-based decisions.
- Semantic enrichment using ontologies adhering to public standards and custom vocabularies.
- Industry-leading disambiguation.
- Relationship data structured with GeneTree ontology and UniProt.
- Therapeutic area data annotated with named entity identifiers.

Flexible delivery

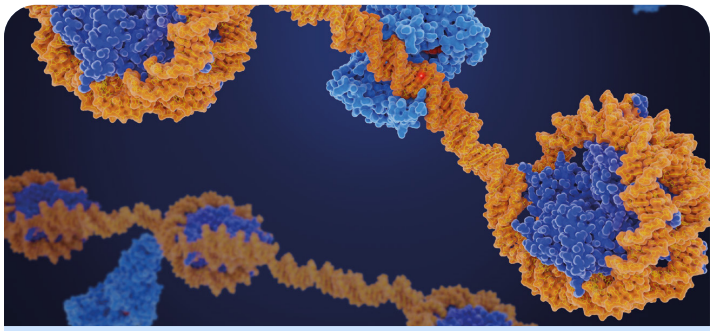
Supporting interoperability with your workflows. Enhanced support (consultancy and scripts) for effective data integration.

API

JSON

Flat files

XML



Case study

Astra Zeneca – driving epigenetic target discovery

Discover how Elsevier data increased the efficiency and depth of target discovery in the epigenetic space.

Read the full story

"EpiMap [built from Elsevier datasets] *was critical in driving novel hypothesis generation"*

Director, Oncology Data Science, AstraZeneca

"I want to help power my decisions using the Embiology [target/relationship] *data"*

Director, Informatics & Data, European Pharmaceutical Company

Improve your data foundations – learn more: elsevier.com/solutions/datasets