Elucidata

Elucidata
Data Modalities

October, 2025

Elucidata supports **30+ data modalities**, including omics, clinical, and imaging data, to enable scalable data product generation, multi-modal integration, and AI-driven insights. Leveraging both **public and restricted data sources** across omics, imaging, and text-based modalities, Elucidata delivers high-quality, AI-ready data for drug discovery, translational research, and precision medicine.

1. Public Data Sources

The table below lists the public data modalities we support, with a defined pipeline, standard curation, and set turnaround time. However, we are happy to customize the pipeline and curation efforts for these data types.

1.1 Public Tabular Data Sources

Solution	Data Type	Source	Description	Pipeline
Single-cell RNA seq	Omics	GEO	Transcriptomics count matrix and associated cell metadata	CellRanger
		CZI	Single cell data repository	
		SCP	Single cell multi-omics	
		ArrayExpress	Transcriptomics data repository	
		CellXGene	Single cell data repository	

		Expression Atlas	Gene expression data	
Single cell Spatial transcriptomics	Omics	GEO	Spatially resolved transcriptomics data	SpaceRanger
		SRA	Single-cell spatial transcriptomics data	
		CZI	Single cell spatial transcriptomics repository	
		SCP	Single cell spatial transcriptomics multi-omics	
		Single Cell Expression Atlas	Single cell expression data repository	
		Allen Brain Cell Atlas	Brain cell spatial data repository	
Bulk RNA seq	Omics	GEO	Transcriptomics, mutation data repository	Kallisto/STAR/Salmon
		CPTAC	PanCancer multi-omics and clinical data	
		TCGA	PanCancer multi-omics and clinical data	

		GTEx	Gene expression across tissues	
Proteomics	Omics	PRIDE	Proteomics data	MaxQuant
		CPTAC	PanCancer multi-omics and clinical data	
		Human Protein Atlas	Protein expression profiles in tissues/cell lines	
Metabolomics	Omics	TEDDY	Multiomics for Type 1 Diabetes	XCMS, Elmaven
		CPTAC	PanCancer multi-omics and clinical data	
		MetaboLights	Metabolomics experimental and results	
CITESeq	Omics	GEO	Transcriptomics, mutation, single cell data repository	On-demand
		CZI	Single cell data repository	
		SCP	Single cell multi-omics	
scATACSeq	Omics	GEO	Transcriptomics, mutation, single cell data repository	On-demand

		CZI	Single cell data repository	
		SCP	Single cell multi-omics	
WES	Omics	GDC	Cancer genomics data	GATK, Mutect2
		TCGA	PanCancer multi-omics and clinical data	
		cBioPortal	PanCancer multi-omics and clinical data	
		CPTAC	PanCancer multi-omics and clinical data	
WGS	Omics	GDC	Cancer genomics data	GATK
		TCGA	PanCancer multi-omics and clinical data	
		cBioPortal	PanCancer multi-omics and clinical data	
		CPTAC	PanCancer multi-omics and clinical data	
Methylation	Omics	CPTAC	PanCancer multi-omics and clinical data	IlluminaHumanMethylationEPIC Array
		cBioPortal	PanCancer multi-omics and clinical data	

		TCGA	PanCancer multi-omics quantification and open clinical data	
GWAS	Omics	UKBioBank	GWAS summaries	On-demand
		gnomAD	Genetic variant frequency data in populations	
		GWAS Catalog	Genotype-phenotype association summaries	
Gene Dependencies	Assay	DepMap	Gene dependencies and CRISPR screens for cancer cell lines	Not Applicable
Drug Response	Assay	PharmacoDB	Drug response database	Not Applicable
Cytometry	Assay	Immport	Clinical Trial Data, assay data	Not Applicable
		FlowRepository	Flow cytometry data repository	
IHC/Elispot/ELISA/ PCR	Assay	Immport	Clinical Trial Data, assay data	Not Applicable

1.2 Public Imaging Data Sources

Solution	Data Type	Source	Description	Pipeline
HE/HES slides	Imaging	IDC	Cancer imaging data	Not Applicable
		TIGER	Tumor-infiltrating lymphocytes in H&E breast cancer slides	
		BreaKHis	Breast cancer histopathology images	
		TCGA	Cancer imaging data	
MRI	Imaging	IDC	Cancer imaging data	Not Applicable
CT	Imaging	IDC	Cancer imaging data	Not Applicable

1.3 Public Text Data Sources

Source	Description
PubMed	Literature and clinical trial summaries
ClinicalTrial.gov and AACT	Clinical studies conducted around the world, providing detailed information about their design, status, and result

2. Restricted Data Sources

These restricted sources include **tabular, text and imaging** datatypes that require approvals from data providers, which may involve additional compliance steps and impact turnaround times. Some repositories contain both publicly available and restricted data, we can work with the publicly available portion without additional approvals. However, accessing restricted sections will require specific permissions, potentially affecting project timelines.

2.1 Restricted Tabular Data Sources

Solution	Data Type	Source	Description	Pipeline
Raw Sequencing Data or Tabular Data	Omics	dbGaP	Database of Genotypes and Phenotypes to access raw sequencing data from human samples	Varies
		PPMI	Parkinson's Progression Markers Initiative	
		EGA	Genomics and transcriptomics raw sequencing data from human samples	
		UKBiobank	Population health and genetic data	

2.2 Restricted Text and Imaging data sources

Solution	Data Type	Source	Description	Pipeline
Electronic Health Records (EHR) or Real World Evidence (RWE)	Text	Premium Data (Hospitals and RWE Data Providers)	Clinical records and structured health data	Not Applicable
Restricted Imaging	Imaging	TCIA	PET imaging dataset linked to TCIA	Not Applicable
		BraTS	Brain tumor MRI dataset	Not Applicable

3. Additional Resources

3.1 Public Biomedical Databases

Database	Description
NCBI Gene	Database for Gene, sequence and annotation
Uniprot	Database of Protein, sequence and annotation
CHEMBL	Database of bioactive molecules with drug-like properties, containing information on their chemical structures, biological activities, and target interactions

OpenTargets	End points covered • *association_by_datatype_direct - Returns direct associations between targets and diseases, filtered by a specific data type* • *disease_phenotype - Maps diseases to associated phenotypes* • *disease - Provides names and IDs of diseases, including their cross-references to other diseases.* • *drug_indication - Indications from clinical trials* • *interaction - Target pair-wise molecular interactions* • *target - Contains target information and tractability information for each target*
COMPARTMENTS	Subcellular protein localization and interactions
Reactome	Pathway annotations
STRINGdb	Protein interaction database
Diseases DB	Disease-gene associations mined from literature

3.2 Public Biomedical Ontologies

Ontology	Description
GO	Gene Ontology annotations
HGNC	Official gene nomenclature
EFO	Experimental factor ontology
Mondo	Disease ontology

Cell Ontology	Standard cell type classifications
Uberon	Anatomical structure ontology
HPO	Human Phenotype Ontology

***Note**:* We are open to collaborate and provide custom pipeline and curation solutions for all listed modalities.