

BioBricks-OKG: An Open Knowledge Graph For Cheminformatics And Chemical Safety

Thomas Leuchtefeld, Insilica

Use case description and societal challenge being addressed

The BioBricks-OKG is an open knowledge graph focused on chemical safety.

Knowledge graph source datasets

BioBricks-OKG draws on primary sources supplied by biobricks.ai, a list of the sources that may be included is available at <https://biobricks.ai/search>. Today, BioBricks-OKG integrate:

Graph	Triples (m)	Description
HGNC	~ 1	Overview of unique symbols and names for human genes.
ICE	~ 7	Organized toxicological results for endpoints like skin sensitization, eye irritation, etc.
MESH	~ 18	Common medical vocabulary used in ICE & Pubmed
TOXCAST	~ 24	A large set of assays with hitcall results.
TOX21	~ 131	A large set of assays with hitcall results.
UNIPROT-KG	~ 144,142	links proteins to the assays shared in other data sets and provides general protein information.
PUBCHEM-RDF	~ 242,542	A graph of 27 subgraphs related to a variety of chemical information.
TOTAL	~ 384,323	

User queries / competency queries for the use case

- Chemical identifier lookup: Look up the names of chemical entities and optionally their CAS RN and DSSTOXSID if available. Indicate which graph the data come from.
- Assays by source: Look up assays from InvitroDB that are found in both ICE and ToxCast by graph (where the graph IRI corresponds to which dataset it is from).
- Endpoint responses: Retrieves endpoint responses chemicals and assays. An endpoint response provides information about observed experimental outcomes when chemicals are tested in assays. Endpoint responses include the (often numeric) value itself and entities like units (eg. mg/kg).