



- **RTX-KG2** "... is the second-generation knowledge graph for the ARAX biomedical reasoning system."
- Includes a pipeline and data output.
- Some data not publicly available due to licensing.



- RTX-KG2 Versions:
 - **"non-canonicalized"**: untranslated from data sources.
 - **"canonicalized"**: builds a unified new ontology specific to RTX-KG2
 - **"lite"**: removes certain nodes and relationships removed to slim the graph down.



Reproducing RTX-KG2 Pipeline Execution

- "Extended Pipeline" fork on GitHub
- Attempted modified Singularity-based execution on HPC Alpine: likely requires significant redesign.
- Attempting GCP VM-based execution: some progress but blocked by data requirements and existing software bugs.
 - See: [RTX-KG2#366](#), [RTX-KG2#365](#), [RTX-KG2#362](#), [RTX-KG2#303](#) (comment)

Accessing RTX-KG2 Data Exports

- "RTX-KG2-gateway" repo on GitHub
- JSON property graph available at [translator-lfs-artifacts](#) offers one-way use of data (17.63 GB JSON file, 686 MB compressed).
- Used Parquet for intermediary for ingest.

RTX-KG2 Data through Kuzu



- **Kuzu** is an embedded graph database technology.
- **Kuzu database of RTX-KG2** successfully created (23 GB directory, 1 GB compressed).
- Working on support for the project through Cypher queries and exploration of schema.