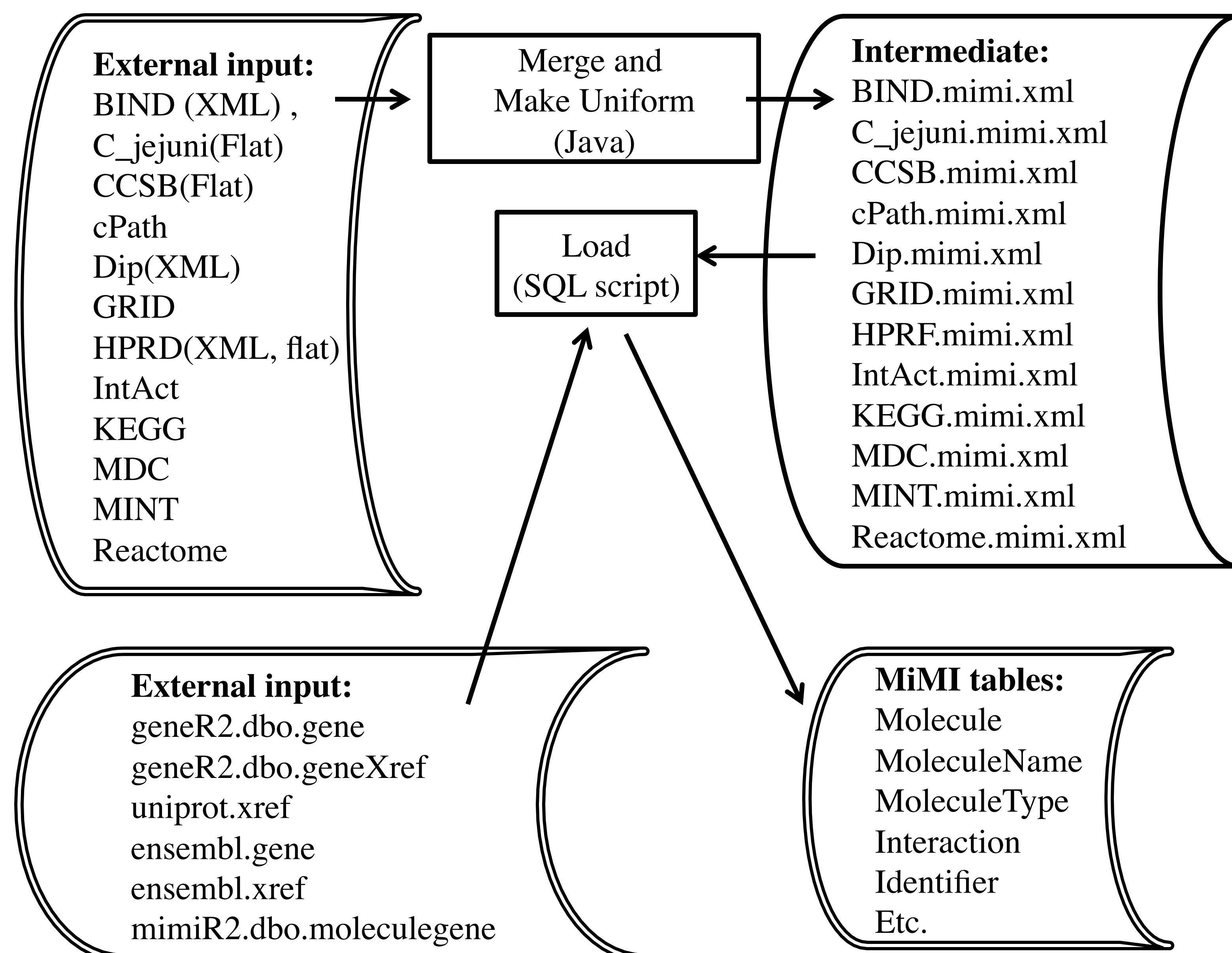


1. Integration Workflow of MiMI



6. Examples of Provenance in Intermediate File

Example 1

('ChEBI', 10003) in Identifier table ⇒

| File name | MoleculeID | MoleculeType | idvalue | idtype |
|----------------------|------------|--------------|---------|--------|
| KEGG.ligand.mimi.xml | C01759 | Compound | 10003 | ChEBI |

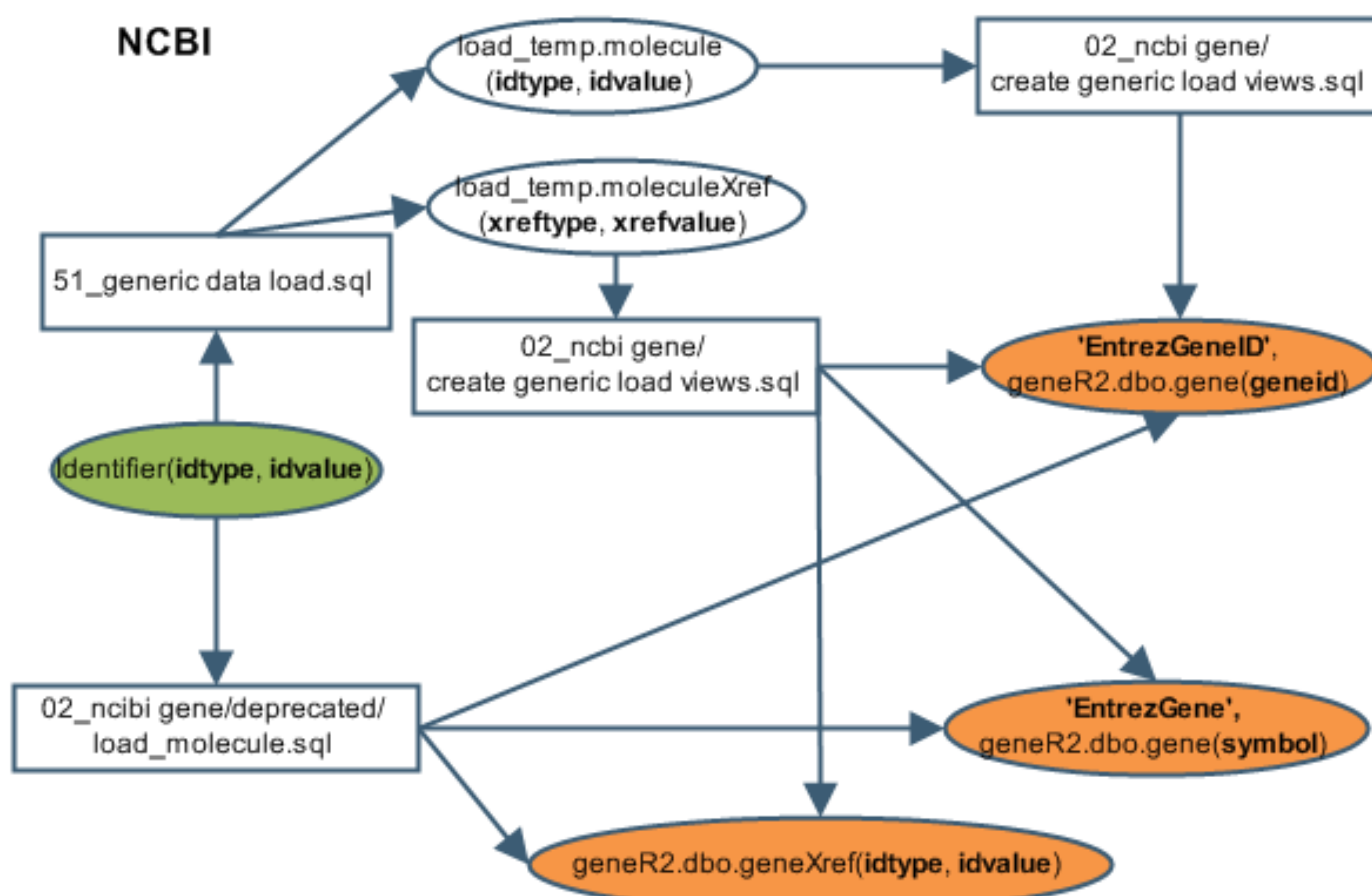
Example 2

('CSF1R' vs. 'INPP5D') in Interaction table ⇒

| File name | Interaction ID | Molecule A | Molecule B | Interaction Type | PubMedID | Description |
|---------------|----------------|------------|------------|------------------|----------|-------------------------------------|
| BIND.mimi.xml | 3493 | 833 | 1026 | Bidirectional | 15215307 | CSF-1 receptor interacts with SHIP. |

Data flows involved in Example 1

Identifier (MiMI table) ⇒ NCBI genes (source database)



2. Integration In Two Steps

- First, external source files ⇒ intermediate XML files
- Second, intermediate XML files ⇒ MiMI tables
- External files are in various formats
- Intermediate XML files are in a uniform internal format

3. Provenance of Tuples in MiMI

- Provenance of a tuple in MiMI includes
 - Contributing XML elements in intermediate files and contributing tuples in gene tables
 - Contributing data items in external source files
 - More difficult to retrieve
 - Merging script is more complicated than loading script
 - External files have different formats and thus different access methods
 - This is next step work

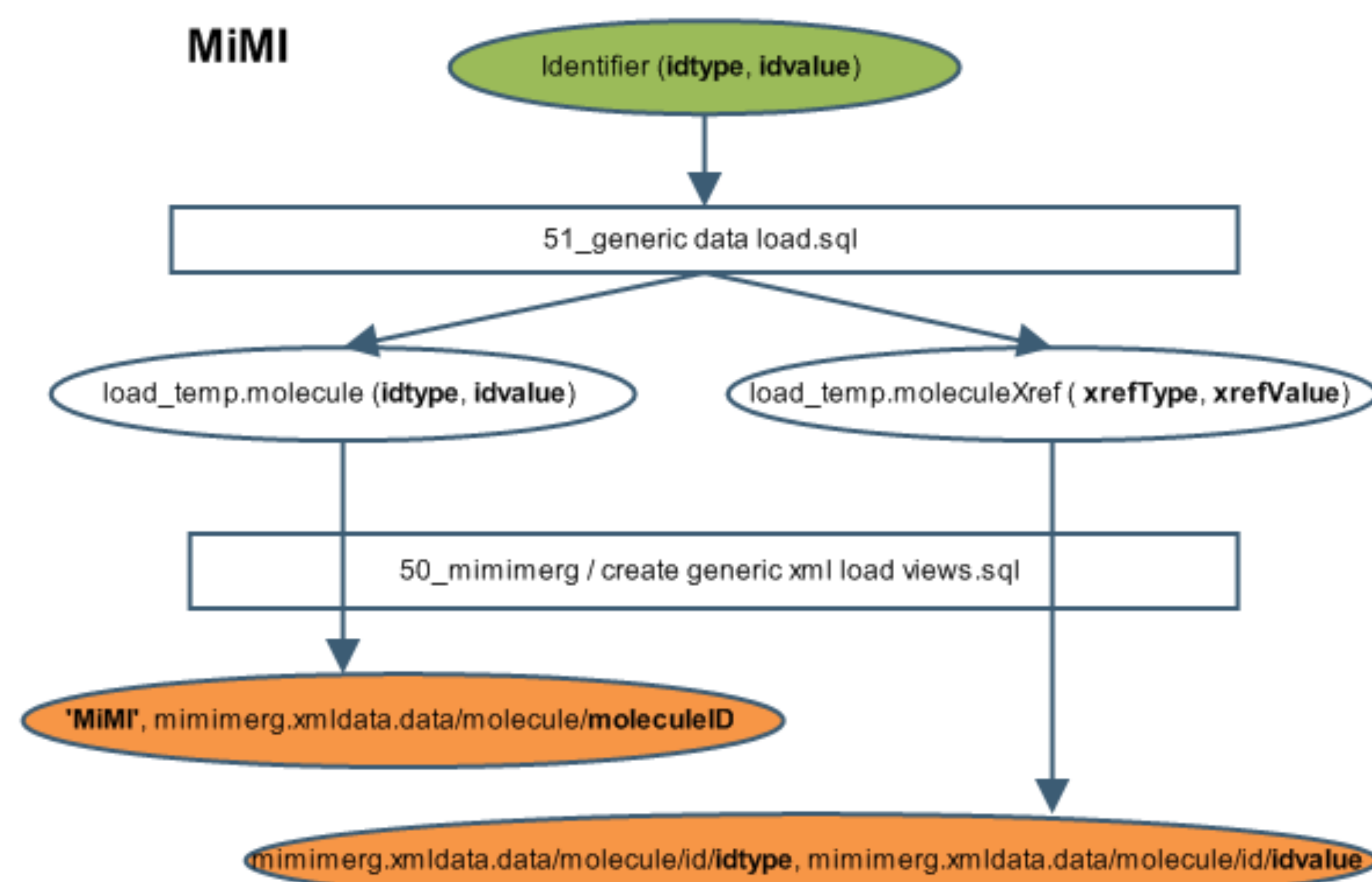
4. Tracing Provenance for MiMI Tuples in Intermediate Files and Gene Tables

- Intermediate files are *.mimi.xml
- For the purpose of fast provenance retrieval, these files are transformed ("flattened") into tables
- Based on the dataflow of the loading script, tracing queries are generated to query the transformed intermediate tables and gene tables

5. Roadmap of Retrieving Provenance in Intermediate Files

- For each MiMI table
 - Figure out the set of contributing source table
 - For each contributing source table
 - Draw a data flow diagram that describe how tuples in this source table travel to the MiMI table
 - Write a tracing query that retrieve provenance from this source table
 - Provenance for tuples in this MiMI table is the union of contributing tuples from all involved source tables

Identifier (MiMI table) ⇒ MiMI XML files



Acknowledgements

This work was supported by National Institutes of Health: Grant #U54 DA021519