

Data graphs have ever more...

...Systems supporting them

- Cosmos DB
- Neo4j
- Spark
- Datastax
- OrientDB
- Virtuoso
- JanusGraph
- Redis

...Languages to query them

- Cypher
- GraphQL
- Gremlin
- G-Core
- PGQL
- SPARQL

...Real-life applications

- Social networks
- Semantic web
- Bioinformatics
- Fraud detection
- Journalism (Panama papers)

Property graphs, data graphs in practice

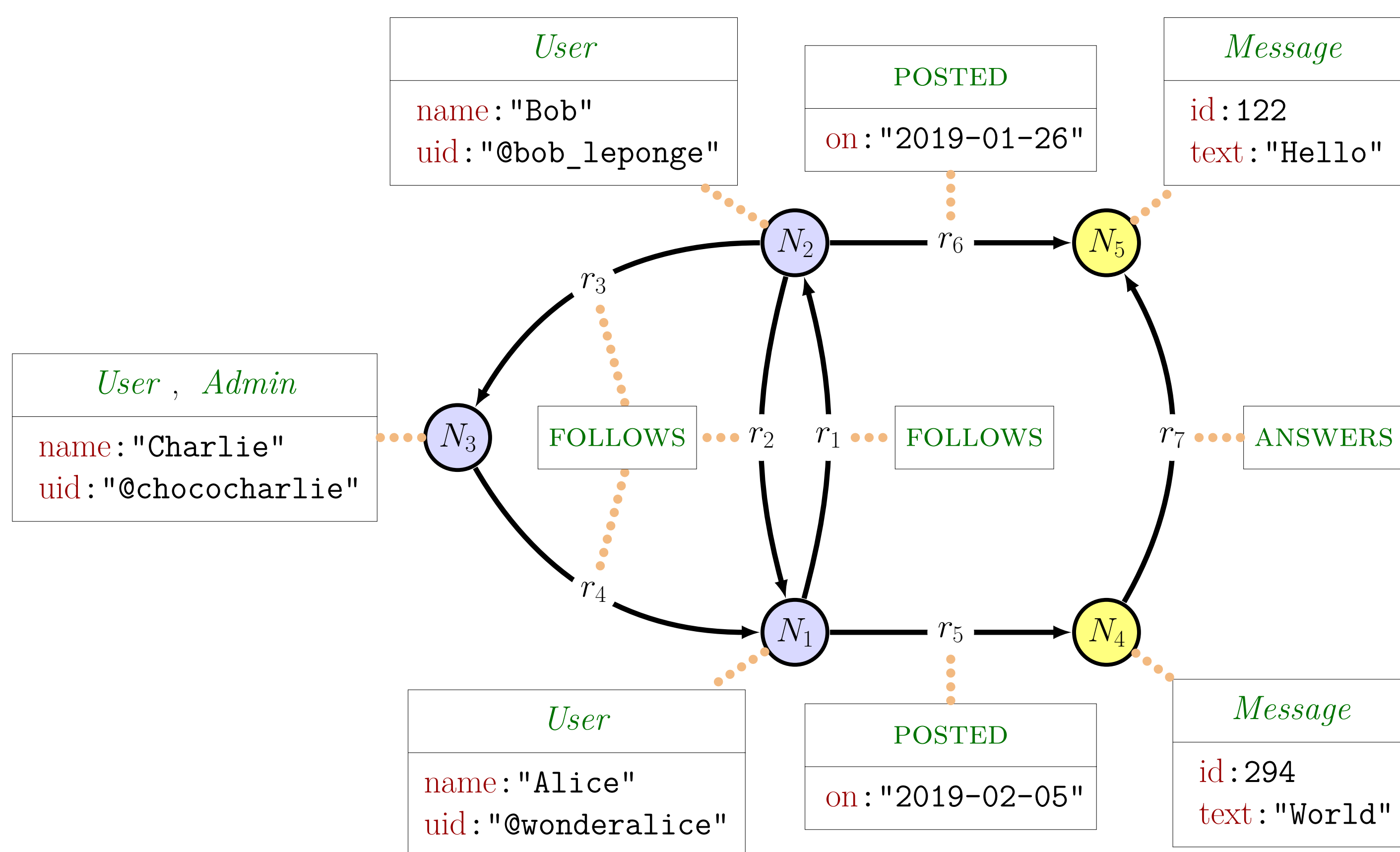


Figure 1: A property graph showing a social network

Reasoning with incomplete data

Data management systems nowadays often have to cope with large volumes of incomplete or inconsistent data. This underlines the need for reasoning tools that can efficiently produce reliable answers from uncertainty. Typical scenarios include data exchange and data leak prevention.

Data exchange for data graphs

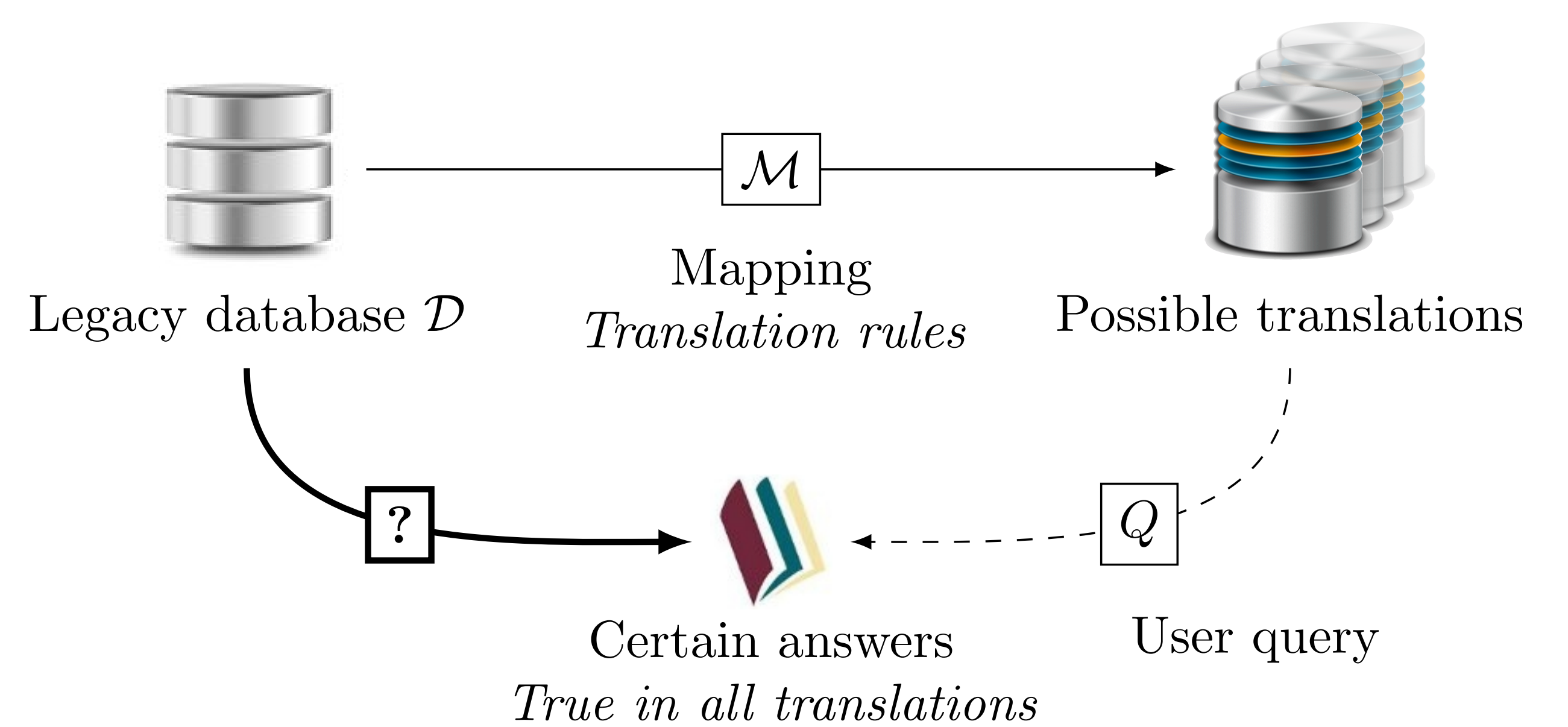


Figure 2: Data exchange scenario. Can we return the certain answers to user queries on the translated schema without computing the (possibly infinite) set of possible answers?

Cypher, leading query language in industry

Example Cypher-query Evaluation

Querying news since 2019, February 1st

- ```
(C1) MATCH (u:User)-[p:POSTED]->(m:Message)
(C2) WHERE p.date >= "2019-02-01"
(C3) MATCH (f:User)-[:FOLLOWS]->(u)
(C4) RETURN f.uid AS i, u.name AS n, m.text AS t
```

#### Result: (Cypher queries return tables)

| receiverId      | senderName | message   |
|-----------------|------------|-----------|
| "@bob_leponge"  | "Alice"    | "I'm in!" |
| "@chococharlie" | "Alice"    | "I'm in!" |

### Theorem [5]

Data complexity of answering a data regular path query is:

- **Undecidable** under mappings with copy and reachability rules.
- **coNP-hard** if the mapping is restricted to rules bounded length; and drops to **NLogSpace** if the data domain contains **null**.

### Formal denotational semantics of Cypher

- Covers all major constructs of the language
  - Expression [1, 2]
  - Core read-only clauses [1, 2]
  - Aggregation and row ordering [2]
  - Update clauses [3, 2]
- A step towards standardisation of Cypher-query evaluation.
- Reference implementation : automated testing of compliance [4]

### Future work: data leak prevention [6]

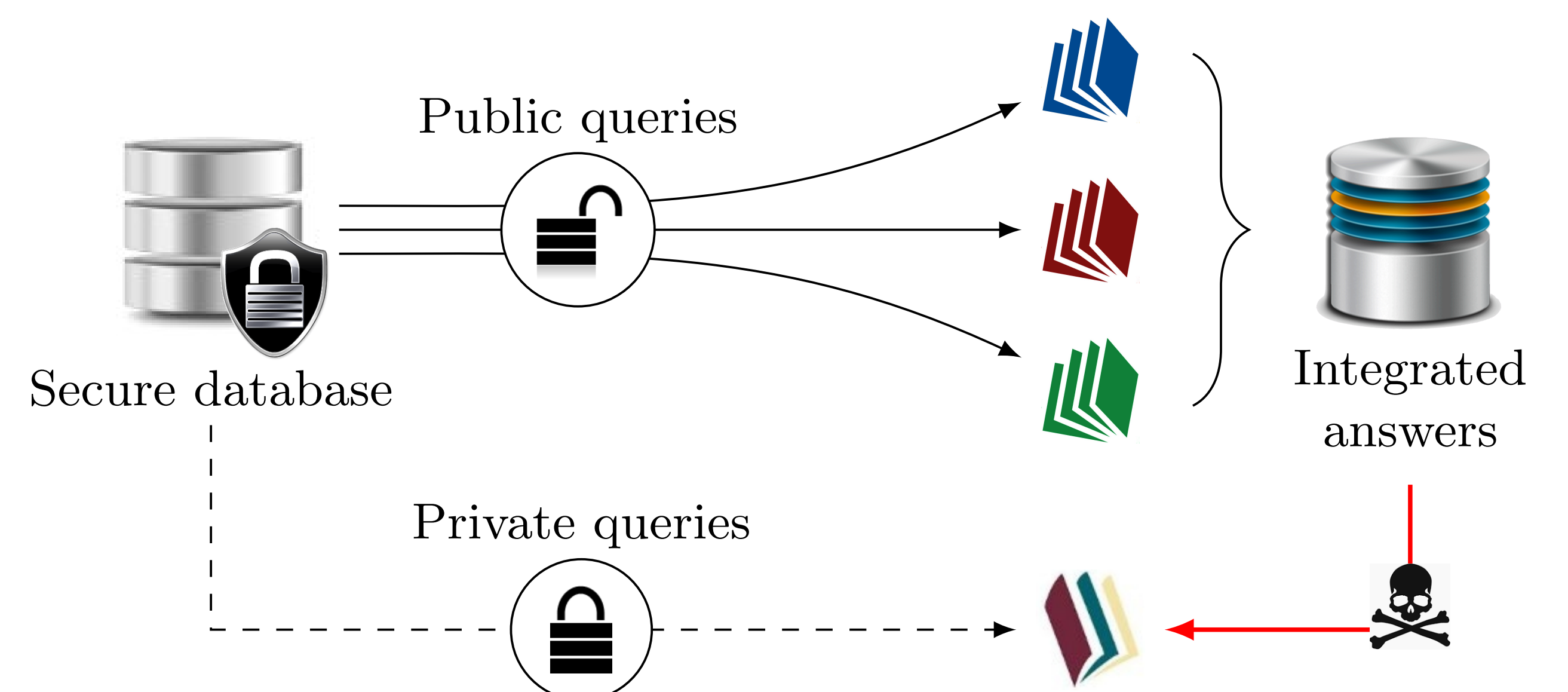


Figure 3: Can we ensure that private data cannot be computed from public answers?

## References

- [1] N. Francis et al. "Cypher: An Evolving Query Language for Property Graphs". In: *Proceedings of SIGMOD'18*. ACM Press, 2018.
- [2] N. Francis et al. *Formal Semantics of the Language Cypher*. Tech. rep. Arxiv:1802.09984. 2018.
- [3] A. Green et al. "Updating Graph Databases with Cypher". Submitted. 2019.
- [4] V. Marsault et al. *CyphSem: Formal Semantics of Cypher for developers*. Java Library.
- [5] N. Francis and L. Libkin. "Schema Mappings for Data Graphs". In: *PODS'17*. ACM Press, May 2017.
- [6] C. David et al. *Efficiently Querying Incomplete and Inconsistent Databases (QUID)*. ANR-18-CE40-0031. 2019.