

# Rethinking Biomedical Reasoning: The Strategic Role of SMT over Traditional Knowledge Graphs

Munongedzi Mabhoko, Clarkson university, [mabhokm@clarkson.edu](mailto:mabhokm@clarkson.edu)

## *Abstract*

Biomedical informatics increasingly relies on structured representations and automated reasoning to support data integration, clinical decision-making, and biomedical discovery. While knowledge graphs (KGs) have dominated this space as a flexible tool for integrating large-scale data, they often fall short in scenarios demanding formal guarantees, semantic precision, and logical safety. This paper argues that Satisfiability Modulo Theories (SMT)-based approaches offer a more rigorous, verifiable, and semantically robust framework, particularly in ontology-based biomedical workflows. Drawing on recent advances in SMT-based safety verification and semantic repair, it demonstrates how SMT systems outperform KGs in tasks involving data-aware processes, inconsistency detection, and protocol verification.

## 1. Introduction

The biomedical domain is marked by growing data complexity, rich ontological hierarchies, and high-stakes decision contexts. Knowledge graphs (KGs), built upon RDF and OWL standards, have been widely used to capture and relate biomedical entities such as genes, drugs, and diseases. While useful for semantic search and integration, KGs typically offer limited inferential power and cannot ensure correctness across evolving ontologies or dynamic processes.

In contrast, **Satisfiability Modulo Theories (SMT)** provides a formal logic-based foundation capable of expressing and verifying complex biomedical rules, constraints, and

behaviors. Recent work demonstrates that SMT can address key gaps in KGs, particularly when applied to data-aware processes, ontology repair, and semantic verification.

## 2. Limitations of Knowledge Graphs in Biomedical Contexts

Knowledge graphs such as **BioKG**, **PathMe**, and **Hetionet** are built by integrating structured biomedical data using entity-relation triples. These graphs are optimized for **semantic linking**, **visualization**, and **querying** via SPARQL or graph search.

However, challenges emerge:

- **Limited reasoning depth:** Most KG inference is limited to transitivity, subclass hierarchies, or OWL DL subset.
- **No safety guarantees:** KGs cannot formally verify that a medical workflow is logically sound or avoid unsafe states.
- **Hidden contradictions:** Combining multiple ontologies often leads to unsatisfiable classes not easily detectable via triple-based systems.

These shortcomings are highlighted in *Slater et al. (2020)*, where the combination of ontologies from the **OBO Foundry** revealed over **300,000 hidden unsatisfiable classes**, largely undetected by standard KG tools (Slater, 2020).

## 3. SMT for Biomedical Safety Verification

In response to these limitations, SMT-based methods offer a mathematically grounded framework to model, verify, and explain complex biomedical logic. In *Calvanese et al. (2021)*, the authors develop a verification technique over **Simple Artifact Systems (SASs)** that operate not on traditional databases but over description logic ontologies. Using SMT

solvers and backward reachability, they formally verify whether data-aware biomedical processes can lead to unsafe states (Calvanese, 2021).

Similarly, the *Formal Modeling and SMT-Based Verification of Data-Aware BPMN* framework enables the encoding of clinical or biomedical workflows with persistent data into SMT formulas. These can be checked using automated tools such as **MCMT**, a state-of-the-art SMT-based model checker.

These systems:

- Provide guarantees about process safety.
- Work with infinite data domains, unlike finite KG traversals.
- Model both data and control logic, critical for biomedical protocol verification.

## **4. Repairing Biomedical Ontologies Using SMT-Inspired Reasoning**

Biomedical ontologies are central to structuring domain knowledge, but as shown by recent research, combining large ontologies often introduces logical contradictions that render parts of the knowledge base unusable. Their work uses reasoning approaches to detect and minimally repair these contradictions, identifying a small set of axioms whose removal resolves unsatisfiability across hundreds of thousands of classes. This kind of reasoning parallels SMT's strengths:

- Explanation of inconsistencies through formal justifications.
- Minimal repair sets for restoring coherence.
- Application of model-theoretic principles to ontology engineering.

Such capabilities are currently beyond the reach of conventional KGs, which lack built-in logical consistency enforcement.

## 5. SMT vs KGs in Biomedicine

Capability	SMT-Based Systems	Knowledge Graphs
<b>Formal Safety Verification</b>	Proven correctness under logical rules	No guarantee beyond triple structure
<b>Reasoning over Infinite Domains</b>	Supports quantified variables and array structures	Limited to finite graph traversals
<b>Semantic Repair and Debugging</b>	Detects and resolves contradictions with justifications	Lacks built-in ontology repair
<b>Scalability</b>	Computationally intensive (PSPACE in worst cases)	Efficient for large-scale query and visualization
<b>Ease of Use and Integration</b>	Requires formal specification, modeling effort	Compatible with standard web and data tools

## 6. Hybrid Outlook The Best of Both Worlds

While SMT-based systems excel in logical precision and formal guarantees, knowledge graphs remain essential for data integration and semantic interoperability. Future biomedical systems could benefit from hybrid architectures, where

- KGs provide the **semantic layer** and entity mapping,
- SMT systems enforce **process constraints** and verify **critical reasoning tasks** (e.g protocol execution, clinical workflow safety).

This approach supports both **scalable data access** and **high-assurance reasoning**, essential for domains like clinical informatics, genomics, and drug interaction modeling.

## 7. Conclusion

SMT-based reasoning is not a replacement for knowledge graphs but a complementary extension that adds logical rigor to biomedical reasoning. As the biomedical field moves toward automated, interpretable, and verifiable systems, SMT methods will be indispensable ensuring trustworthiness in data-driven medical decisions, detecting ontology inconsistencies, and formally verifying process correctness where it matters most.

## References

- Calvanese, D., Gianola, A., Mazzullo, A., & Montali, M. (2021). SMT-Based Safety Verification of Data-Aware Processes under Ontologies. *arXiv preprint*. [arXiv:2108.12330](https://arxiv.org/abs/2108.12330)
- Calvanese, D., Ghilardi, S., Gianola, A., Montali, M., & Rivkin, A. (2019). Formal Modeling and SMT-Based Parameterized Verification of Data-Aware BPMN (Extended Version). *arXiv preprint*. [arXiv:1906.07811](https://arxiv.org/abs/1906.07811)
- Slater, L. T., Gkoutos, G. V., & Hoehndorf, R. (2020). Towards semantic interoperability: finding and repairing hidden contradictions in biomedical ontologies. *BMC Medical Informatics and Decision Making*, 20(1), 1–15. [DOI: 10.1186/s12911-020-01336-2](https://doi.org/10.1186/s12911-020-01336-2)