American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX • Website: www.ajamir.com • Email: editor@ajamir.com

Semantic Knowledge Graphs for Integrating Fragmented Institutional Evidence

Dr. Ethan Marshall

Associate Professor, University of Amsterdam, Netherlands

Abstract

Institutional evidence is frequently dispersed across policy documents, research reports, operational databases, spreadsheets, administrative records, project repositories, technical systems, archival collections, and departmental knowledge bases. Although these sources may individually contain valuable information, differences in terminology, identifiers, schemas, formats, provenance, access conditions, and update cycles can prevent institutions from constructing a coherent evidential view. Semantic knowledge graphs provide a promising approach by representing entities, concepts, claims, sources, and their relationships within a machine-interpretable graph while preserving links to originating evidence. This study develops a Semantic Institutional Evidence Integration Framework (SIEIF) combining ontology-mediated integration, entity resolution, semantic mapping, provenance modeling, graph validation, contradiction preservation, access governance, and evidence-aware retrieval. The framework draws upon contemporary Semantic Web infrastructure, including RDF, OWL, SPARQL, SHACL, and PROV-O. W3C's 2026 RDF 1.2 and SPARQL 1.2 technical specifications further reinforce the continuing development of graph-based semantic data exchange and cross-source querying. Because no verified institutional deployment dataset was available, the study adopts an explicitly simulation-based methodology.

Four hypothetical integration architectures are compared across six dimensions: semantic alignment, entity resolution, provenance traceability, cross-source retrieval, contradiction handling, and governance observability. The simulated Semantic Evidence Integration Readiness Index increases from 33 for a fragmented evidence environment to 89 for a provenance-aware semantic knowledge graph. Recent cross-institutional biomedical implementations demonstrate that RDF-based knowledge-graph pipelines can support semantic interoperability across heterogeneous source systems while maintaining local governance and validation. The study concludes that institutional knowledge graphs should not merely collapse heterogeneous evidence into unified facts. Stronger systems preserve provenance, uncertainty, disagreement, version history, and access rights so that integrated evidence remains interpretable, auditable, and contestable.

Keywords: semantic knowledge graphs, institutional evidence, semantic interoperability, ontology, provenance, entity resolution, evidence integration, RDF, knowledge management

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

1. Introduction

Modern institutions generate evidence through a wide variety of organizational processes. Universities maintain research repositories, committee records, policies, project files, student and administrative systems; healthcare organizations manage clinical, operational, research, and regulatory information; public agencies combine legislation, service records, evaluations, correspondence, and administrative databases; and corporations operate across document-management systems, databases, technical repositories, and specialized departmental platforms. The central information problem is therefore rarely a complete absence of data. More often, relevant evidence exists but remains fragmented across incompatible systems. Knowledge-graph research has become increasingly important in precisely these circumstances because graph-based representations can integrate diverse and dynamic information while explicitly modeling relations among entities. Hogan and colleagues' comprehensive review identifies data integration, querying, reasoning, validation, and graph management as central elements of modern knowledge-graph systems.

Fragmentation is both technical and semantic. Two departments may refer to the same person by different identifiers, use different labels for the same organizational process, represent dates or statuses differently, or attach different meanings to apparently identical terms. A traditional integration process can move these records into one repository without resolving their conceptual differences. Semantic integration addresses this problem by introducing shared vocabularies, ontologies, mappings, and machine-readable relationships. RDF is explicitly designed as a model for data interchange and facilitates the merging of data even when underlying schemas differ, while OWL supports formally defined representation of classes, properties, entities, and relationships. Virtual Knowledge Graph and ontology-based data-access approaches provide an alternative to physically copying every source into a new store by placing an ontology-mediated semantic layer over existing systems. Xiao and colleagues describe this paradigm as a means of integrating and accessing heterogeneous data through a unified conceptual view.

Evidence integration requires more than vocabulary alignment because institutional decisions depend upon where information originated and how it was produced. A statement contained in an approved policy, an unpublished draft, an automated report, an expert assessment, and an external research publication cannot necessarily be treated as evidentially equivalent. W3C PROV-O was designed to represent and exchange provenance across different systems and contexts, allowing entities, activities, and agents associated with information production to be represented explicitly. Recent knowledge-graph research increasingly extends this principle from source lineage toward evidence-aware representation. Zong and colleagues' 2026 EvidenceNet framework, for example, represents experimentally grounded findings as explicit evidence objects so that study context, provenance, and quantitative support are not lost when literature is converted into graph structures.

Institutional evidence may also contain genuine disagreement. One evaluation may support a policy intervention while another questions its effectiveness; an older procedure may conflict with a revised standard; two departments may maintain inconsistent classifications; or a research report may express uncertainty rather than a definitive conclusion. Integrating these sources by replacing competing assertions with a single supposedly authoritative edge can conceal important evidential structure. Menotti and colleagues' provenance-driven nanopublication model explicitly represents supporting and conflicting evidence for multi-source assertions and demonstrates why provenance must remain attached to claims when multiple sources contribute to institutional knowledge. Contemporary work on provenance-enhanced statements similarly argues that attributed claims should not automatically be collapsed into one factual commitment when sources differ in epistemic stance.

2. Objectives of the Study

2.1 To Develop a Semantic Institutional Evidence Integration Framework

The first objective is to develop a systematic framework for integrating heterogeneous institutional evidence through semantic technologies. The proposed framework combines ontology design, RDF-based representation, schema mapping, entity resolution, provenance, validation, and evidence-aware querying so that evidence from different organizational systems can be connected without erasing its original context.

2.2 To Evaluate Integration Quality Beyond Simple Data Consolidation

The second objective is to compare alternative institutional evidence architectures across semantic alignment, entity resolution, provenance traceability, cross-source retrieval, contradiction handling, and governance observability. The study distinguishes physical aggregation from genuine semantic integration because placing records in the same repository does not necessarily make their meanings interoperable. Contemporary semantic-data-integration research similarly identifies semantic layers and knowledge graphs as increasingly important for identifying complex relationships across heterogeneous datasets.

2.3 To Establish Governance Principles for Evidence-Aware Knowledge Graphs

The third objective is to identify controls required for trustworthy institutional use. These include authoritative source registration, provenance capture, graph validation, access control, version management, conflicting-evidence preservation, human review, and transparent distinction between source evidence and machine-generated inference.

3. Review of Literature

3.1 Semantic Integration, Ontologies, and Cross-Silo Knowledge

Knowledge graphs have emerged from the convergence of databases, knowledge representation, Semantic Web technologies, information extraction, and graph analytics. Hogan and colleagues describe knowledge graphs as structures that can represent diverse entities and relationships while supporting querying, validation, reasoning, and other downstream applications. Their importance to fragmented institutional evidence lies in the ability to represent heterogeneous information according to relationships that remain explicit rather than hidden within application-specific tables or document structures.

Semantic integration provides the conceptual foundation for this approach. Masmoudi and colleagues' 2022 survey of semantic data integration and querying characterizes the field as one concerned with heterogeneous data whose meaning must be reconciled before meaningful integrated querying becomes possible. Ontologies provide formal vocabularies through which institutional concepts can be represented consistently. OWL supports machine-interpretable definitions of classes and properties, while SKOS provides a common model for taxonomies, thesauri, classification systems, and related knowledge-organization structures. An institution can therefore preserve local terminology while mapping terms to shared concepts rather than imposing one rigid label across every source system.

RDF provides a graph-oriented representation based on subject–predicate–object statements. W3C's RDF 1.2 technical specification, published in April 2026, continues this model while supporting richer graph representation features. SPARQL provides the corresponding query layer and can express graph-pattern queries across diverse data sources, including sources exposed as RDF through middleware. This combination is particularly useful when institutional questions cross organizational boundaries, such as identifying which policies apply to a project, which evidence supports a decision, which organizational units contributed to an initiative, or where contradictory assessments exist.

.

3.2 Entity Resolution, Provenance, and Evidence Quality

Semantic alignment cannot succeed if the same real-world entity is repeatedly represented as unrelated graph nodes. Entity resolution therefore constitutes a core integration task. Zeng and colleagues define entity alignment as identifying entities in different knowledge graphs that refer to the same real-world object and review a range of structural and representation-learning techniques for this purpose. Hofer and colleagues similarly identify entity linking, entity resolution, entity fusion, ontology matching, metadata management, and quality assurance among the necessary components of generalized knowledge-graph construction pipelines.

Institutional evidence makes this challenge particularly acute because identity is frequently context dependent. A researcher may appear in a grant system under one identifier, in publications under another, and in a personnel system under a third. A project may change name while retaining organizational continuity. A department may be reorganized while legacy records preserve its former name. Automatic merging without reliable identifiers can therefore create false equivalence, whereas failure to resolve equivalent entities fragments the evidential picture. Entity resolution must accordingly combine identifiers, attributes, relationships, temporal information, and human review where ambiguity remains.

Provenance provides a complementary layer by explaining where graph statements came from and how they were produced. PROV-O offers a domain-independent foundation for representing provenance information across different systems and contexts. The provenance of a graph assertion may include its originating document, responsible organizational unit, extraction procedure, transformation process, version, timestamp, and reviewing authority. This information becomes critical when an institution needs to audit how an answer was produced or determine whether a relationship reflects a primary source, a derived transformation, or an automated inference.

3.3 Validation, Evidence Retrieval, and Research Gap

High-quality semantic integration requires systematic validation. SHACL provides a formal mechanism for defining constraints on RDF graph structure, including expected classes, properties, and node relationships. W3C's SHACL 1.2 Core technical report published in June 2026 continues this constraint-based validation model. Graph validation can identify missing mandatory properties, invalid identifiers, datatype inconsistencies, incompatible values, and structural errors before problematic evidence becomes available for downstream reasoning.

The practical importance of this approach is demonstrated by the 2026 SPHN Connector. Touré and colleagues describe a schema-driven pipeline that transforms heterogeneous CSV, Excel, JSON, and relational data into semantically enriched RDF and validates results using SHACL and SPARQL. The system has been deployed across Swiss healthcare institutions and addresses semantic interoperability, identifier uniqueness, update management, and local data governance. The implementation is especially instructive because it shows that semantic integration requires expert-driven upstream mapping as well as automated downstream transformation; graph technology does not eliminate the need for domain interpretation.

4. Research Methodology

4.1 Research Design and Synthetic Institutional Evidence Corpus

The study adopts an explicitly simulation-based methodological design. No confidential institutional databases, personnel files, policy repositories, healthcare records, unpublished research archives, or proprietary enterprise systems were used to generate the quantitative analysis.

A synthetic corpus of 600 hypothetical institutional evidence records was conceptually distributed across six source categories: policy and governance documents, research reports, operational databases, project records, technical documentation, and administrative evidence. These evidence classes were selected to reproduce common forms of institutional fragmentation without claiming to model any specific organization.

Four hypothetical integration architectures were compared: a fragmented evidence environment, a federated keyword-search layer, ontology-mediated integration, and a provenance-aware semantic knowledge graph.

Table 1: Semantic Institutional Evidence Integration Framework

Evaluation Dimension	Operational Meaning	Illustrative Technical/Governance Mechanism	Scale
Semantic Alignment	Ability to reconcile heterogeneous schemas, terminology, concepts, and classifications	Ontologies, RDF/RDFS, OWL, SKOS, semantic mappings	0–100
Entity Resolution	Ability to identify equivalent people, projects, organizations, documents, and concepts across sources	Persistent identifiers, attribute matching, relationship-based resolution, human validation	0–100
Provenance Traceability	Ability to trace graph assertions to sources, transformations, agents, and versions	PROV-O, named graphs, source identifiers, timestamps, lineage records	0–100

Evaluation Dimension	Operational Meaning	Illustrative Technical/Governance Mechanism	Scale
Cross-Source Retrieval	Ability to answer questions requiring evidence from several repositories or systems	SPARQL, VKG/OBDA, graph traversal, evidence-aware RAG	0–100
Contradiction Handling	Ability to preserve and expose competing, superseded, uncertain, or conflicting claims	Evidence nodes, assertion status, temporal validity, supporting/conflicting edges	0–100
Governance Observability	Ability to audit access, transformations, validation, graph updates, and generated outputs	SHACL validation, logs, access controls, versioning, review workflows	0–100

Source: Author-developed simulation framework informed by W3C Semantic Web standards, knowledge-graph construction literature, semantic data-integration research, and contemporary evidence-aware graph systems.

4.2 Semantic Integration Pipeline

The proposed integration process begins with source registration, in which each institutional system or document collection receives a persistent source identity and metadata describing ownership, authority, sensitivity, update frequency, and permitted use. This step ensures that evidence governance begins before graph conversion.

The second step is semantic mapping. Local data fields, document concepts, taxonomies, and classifications are mapped to a shared institutional ontology or controlled vocabulary. Not every local distinction should necessarily be eliminated. Where different departments use genuinely different concepts, the graph can preserve both while representing their relationship through broader, narrower, equivalent, or context-specific mappings.

The third step is entity resolution and canonicalization. Candidate representations of the same real-world entity are compared using identifiers, attributes, graph relationships, and temporal context. Ambiguous cases remain unresolved or enter human review rather than being automatically merged.

The fourth step is evidence graph construction, during which claims are connected to their sources rather than transformed directly into decontextualized facts. Provenance records indicate the originating artifact, responsible agent, transformation process, version, and time where available. PROV-O provides a standardized foundation for representing such lineage.

The fifth step is validation and controlled publication. SHACL constraints test structural integrity, while terminology and identifier checks evaluate semantic consistency. The SPHN Connector offers a current practical example of schema-driven RDF transformation combined with SHACL/SPARQL validation and traceability.

4.3 Simulation Scoring and Future Validation

A Semantic Evidence Integration Readiness Index (SEIRI) was calculated on a 0–100 scale using equal conceptual weights across the six dimensions. Equal weighting was chosen for methodological transparency and should not be treated as an empirically validated institutional standard.

The four architectures were defined as follows. The Fragmented Evidence Environment retains evidence in source-specific systems with limited semantic mapping or provenance-aware cross-search. The Federated Keyword Layer provides common search without resolving deeper conceptual differences among sources. The Ontology-Mediated Integration architecture introduces shared semantics, mappings, and stronger entity integration. The Provenance-Aware Semantic Knowledge Graph additionally preserves evidence lineage, contradiction, validation, access governance, and source-grounded retrieval.

Future empirical validation should compare these architectures using real institutional tasks such as cross-repository evidence discovery, duplicate-entity identification, policy provenance tracing, superseded-document detection, conflicting-evidence discovery, and natural-language question answering. Appropriate metrics would include entity-resolution precision and recall, mapping accuracy, graph-constraint violation rate, provenance completeness, query coverage, evidence-source precision, contradiction-detection performance, answer traceability, and human-review time.

5. Analysis

5.1 Comparative Evidence Integration Performance

The simulation illustrates a substantial difference between simply making evidence searchable and constructing a semantically governed knowledge layer.

Table 2: Simulated Institutional Evidence Integration Scores

Architecture	Semantic Alignment	Entity Resolution	Provenance Traceability	Cross-Source Retrieval	Contradiction Handling	Governance Observability	SEIRI
Fragmented Evidence Environment	34	41	29	37	24	32	33
Federated Keyword Search	48	52	43	66	31	46	48
Ontology-Mediated Integration	81	78	67	82	59	72	73

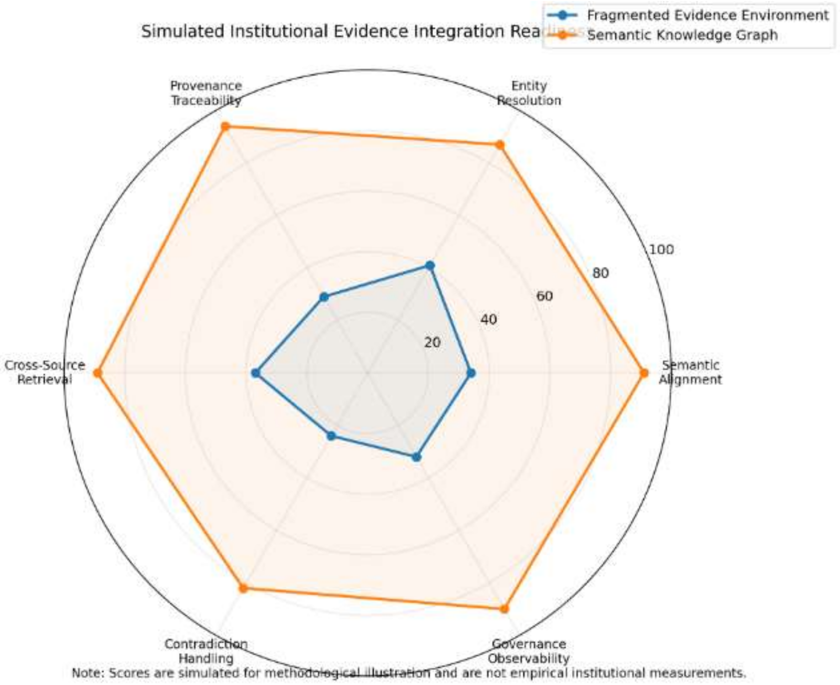
Source: Simulated methodological values; these values do not represent measurements from an actual institution.

The federated search layer produces the strongest relative improvement in cross-source retrieval but remains weak in contradiction handling because discovering documents does not automatically establish how their claims relate.

Ontology-mediated integration substantially improves semantic alignment and entity resolution because evidence can be interpreted through a shared conceptual layer. This aligns with both Virtual Knowledge Graph research and current enterprise ontology-based integration practice.The strongest architecture produces its largest gains in provenance traceability and governance observability. This reflects the paper's central proposition that evidence integration should preserve source accountability rather than simply maximize graph connectivity.

5.2 Institutional Evidence Integration Readiness

Graph 1.:Simulated Institutional Evidence Integration Readiness



The radar graph compares the fragmented evidence environment with the provenance-aware semantic knowledge graph across the six analytical dimensions. The strongest improvement occurs in provenance traceability, increasing from a simulated value of 29 to 94, followed by semantic alignment, governance observability, and cross-source retrieval.

Contradiction handling remains the lowest-scoring dimension even in the strongest architecture, at 82. This reflects an important methodological reality: identifying that two evidence items differ is easier than determining whether they are genuinely contradictory. Differences may result from time, population, methodology, organizational context, terminology, or version changes rather than substantive disagreement.Recent provenance-oriented research supports retaining supporting and conflicting evidence explicitly rather than forcing premature factual consolidation.

5.3 Evidence-Aware Querying and Institutional Decision Support

A semantic graph enables institutional users to pose questions at a conceptual level rather than through source-specific terminology. SPARQL provides formal cross-graph querying, while ontology-mediated systems can translate high-level concepts into underlying data-access operations.

For example, a conventional institutional search system may return every document containing the phrase “research integrity.” A semantic evidence query could instead identify the current policy, its superseded predecessor, committees responsible for revision, supporting consultation records, implementation guidance, relevant training documents, and later audits evaluating compliance.

The distinction becomes increasingly important when natural-language AI interfaces are added. A graph-grounded system can retrieve evidence before generating an explanation and can expose source provenance alongside the answer. FAIR GraphRAG and EvidenceNet provide recent examples of graph-based, evidence-aware retrieval designs that preserve richer metadata and source context than unstructured retrieval alone.

Institutional AI should nevertheless avoid translating graph connectivity directly into authoritative conclusions. The graph may contain outdated, contested, incomplete, or lower-authority evidence. Retrieval therefore needs source ranking, temporal status, provenance visibility, and human review for consequential decisions.

6. Discussion

6.1 Semantic Integration Should Preserve Institutional Meaning Rather Than Merely Combine Records

The central implication of the analysis is that institutional evidence integration should not be evaluated primarily by the quantity of information placed into a common repository. Physical consolidation can reduce technical fragmentation while leaving semantic fragmentation unresolved. Two databases can share a storage platform and still interpret a field differently, while two documents can use identical words to express different institutional concepts. Semantic knowledge graphs provide value because they make conceptual relationships explicit and permit local evidence to be interpreted through a shared but extensible vocabulary. RDF supports graph-based interchange across differing schemas, while OWL and related semantic models enable richer representation of institutional concepts and relationships.

This principle has practical consequences for ontology design. Institutions should resist the temptation to produce one excessively simplified enterprise vocabulary that removes every local distinction. Different professional and administrative contexts may legitimately require different conceptual granularity. The objective should be semantic mediation rather than forced linguistic uniformity. A locally meaningful term can remain within the graph while being connected to a broader institutional concept through an explicit mapping. SKOS and OWL provide different mechanisms for representing such semantic relationships.

The same principle applies to legacy information systems. Replacing every database with a native graph store is neither necessary nor always desirable. Virtual Knowledge Graph approaches demonstrate that heterogeneous data can remain within established systems while being accessed through an ontology-mediated semantic layer. The Novo Nordisk case further illustrates the organizational relevance of this approach where legacy and federated data sources must remain integrated with existing security and governance controls. Institutional semantic integration should therefore be designed around the operational characteristics of source systems rather than around a requirement that every dataset be physically materialized into one graph.

6.2 Provenance and Contradiction Are Core Components of Evidence Quality

The second major implication is that institutional knowledge graphs should distinguish knowledge representation from evidence representation. A graph optimized only for compact facts may state that a policy "improves compliance," that a project "achieved its objective," or that two concepts "are equivalent." Such statements may be useful for retrieval, but their institutional significance depends upon who made the claim, which evidence supported it, when it was valid, and whether competing evidence exists. PROV-O provides a standardized basis for representing lineage across agents, activities, and entities, but institutionally mature evidence graphs should connect this lineage directly with the assertions users rely upon.

Recent research provides useful models for this richer representation. EvidenceNet retains study context and quantitative evidence instead of compressing literature into decontextualized triples, while provenance-driven nanopublications explicitly distinguish supporting from conflicting evidence contributing to an assertion. These approaches are relevant far beyond biomedical literature. Institutional reports, audit findings, committee recommendations, policy evaluations, and technical assessments frequently differ because they answer different questions or use different methods. A graph that preserves those distinctions is more useful than one that presents a synthetic consensus without evidential explanation.

Contradiction should consequently be treated as data rather than automatically as an error. Some inconsistencies genuinely require correction—for example, two identifiers incorrectly assigned to one entity or an invalid date. Other disagreements reflect the evolution of institutional understanding. A knowledge graph should be capable of indicating that one policy superseded another, that one report disputed a previous finding, or that evidence remains uncertain. Contemporary provenance-enhanced statement research specifically examines how attributed claims can coexist without automatically collapsing disagreement into logical inconsistency.

6.3 Governance and Human Validation Remain Essential in Semantic Knowledge Systems

The third implication is that semantic automation cannot remove the need for institutional stewardship. Ontology mapping, entity resolution, source ranking, relationship extraction, and graph construction increasingly benefit from machine-learning and language-model techniques, but all of these processes can introduce errors. Hofer and colleagues identify metadata management, ontology development, canonicalization, data integration, and quality assurance as continuing challenges in generalized knowledge-graph construction. The SPHN Connector similarly automates substantial portions of RDF transformation and validation while explicitly retaining expert responsibility for upstream semantic mapping and source-data accuracy.

Institutions should therefore establish ownership of the semantic layer. Domain experts should govern definitions and mappings; data stewards should oversee source quality and identifiers; information-governance teams should define access rules; technical teams should maintain pipelines and validation; and evidence owners should retain responsibility for the authoritative status of their source material. These responsibilities should remain visible through graph metadata and audit records.

Validation should operate at several levels. Structural validation can use SHACL constraints to identify malformed or incomplete graph patterns. Semantic validation requires domain expertise to determine whether mappings are conceptually appropriate. Entity validation requires review where identity matches remain uncertain. Evidence validation requires confirmation that the graph has represented source claims accurately. Finally, retrieval validation is necessary when a graph is used as grounding for generative AI because semantically related evidence is not always decision-relevant evidence.

7. Conclusion

Fragmented institutional evidence is not merely a storage problem. It is a problem of incompatible meanings, duplicated identities, disconnected sources, uncertain provenance, conflicting claims, changing versions, and uneven access conditions. The present study has developed a Semantic Institutional Evidence Integration Framework that addresses these problems through semantic alignment, entity resolution, provenance traceability, cross-source retrieval, contradiction handling, and governance observability. By connecting institutional records through explicit semantics rather than relying exclusively on physical consolidation or keyword search, knowledge graphs can make relationships among previously isolated sources more visible and computationally accessible.

The simulation-based analysis produced a Semantic Evidence Integration Readiness Index of 33 for fragmented evidence, 48 for federated keyword search, 73 for ontology-mediated integration, and 89 for a provenance-aware semantic knowledge graph. These values are methodological illustrations rather than measurements from an actual organization, but they demonstrate an important conceptual progression. Search improves discovery without necessarily improving meaning; ontology-mediated integration aligns concepts and entities; and provenance-aware knowledge graphs provide a stronger foundation for evidence traceability, contradiction management, validation, and accountable retrieval. The greatest value therefore emerges when semantic integration is accompanied by evidence governance rather than implemented solely as a technical graph-construction exercise.

The study further concludes that institutional knowledge graphs should resist the pressure to transform every source into a single definitive version of truth. Institutions routinely operate with evidence that changes, conflicts, or carries different levels of authority. A trustworthy graph should preserve these distinctions. Provenance should identify where claims originated and how they were transformed; temporal metadata should distinguish current from superseded evidence; contradiction relationships should preserve legitimate disagreement; and access controls should remain aligned with the governance of underlying sources. Recent evidence-aware and provenance-oriented graph research reinforces the value of keeping context and supporting evidence attached to graph assertions.

References

1. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American, 284*(5), 34–43.
2. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data—The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22. https://doi.org/10.4018/jswis.2009081901
3. Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool.
4. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*. https://doi.org/10.1145/3447772
5. Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2020). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2021.3070843
6. Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). *Knowledge Graphs: Methodology, Tools and Selected Use Cases*. Springer. https://doi.org/10.1007/978-3-030-37439-6
7. Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web, 8*(3), 489–508. https://doi.org/10.3233/SW-160218
8. Hogan, A., Harth, A., & Polleres, A. (2007). Scalable authoritative OWL reasoning for the Web. *International Journal on Semantic Web and Information Systems, 3*(3), 1–35.
9. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., & Ngonga Ngomo, A.-C. (2015). DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web, 6*(2), 167–195. https://doi.org/10.3233/SW-140134
10. Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). YAGO: A core of semantic knowledge. *Proceedings of the 16th International Conference on World Wide Web*, 697–706. https://doi.org/10.1145/1242572.1242667
11. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A collaboratively created graph database for structuring human knowledge. *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 1247–1250. https://doi.org/10.1145/1376616.1376746
12. Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *ACM SIGMOD Record, 34*(3), 31–36. https://doi.org/10.1145/1084805.1084812
13. Moreau, L., Freire, J., Futrelle, J., McGrath, R. E., Myers, J., & Paulson, P. (2008). The Open Provenance Model: An overview. *Lecture Notes in Computer Science, 5272*, 323–326. https://doi.org/10.1007/978-3-540-89965-5_31
14. Moreau, L., & Missier, P. (Eds.). (2013). *PROV-DM: The PROV Data Model*. W3C Recommendation.
15. Euzenat, J., & Shvaiko, P. (2013). *Ontology Matching* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-43406-4

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

16. Noy, N. F., & Musen, M. A. (2000). PROMPT: Algorithm and tool for automated ontology merging and alignment. *Proceedings of the 17th National Conference on Artificial Intelligence*, 450–455.
17. Doan, A., Halevy, A. Y., & Ives, Z. G. (2012). *Principles of Data Integration*. Morgan Kaufmann.
18. Halevy, A. Y. (2005). Why your data won't mix. *ACM Queue, 3*(8), 50–58. https://doi.org/10.1145/1103823.1103836
19. Sikos, L. F., & Philp, D. (2020). Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. *Data Science and Engineering*, *5*, 293–316. https://doi.org/10.1007/s41019-020-00118-0
20. Espinoza-Arias, P., Garijo, D., & Corcho, O. (2021). Crossing the chasm between ontology engineering and application development: A survey. *Journal of Web Semantics, 70*, 100655. https://doi.org/10.1016/j.websem.2021.100655