

Towards Provenance-Centric Spatial Data Supply Chains: A Review of Semantic Web Technologies

Muhammad Azeem Sadiq¹, Philip Kibet Langat^{2*}, Arjun Neupane³

^{1,2,3} School of Engineering and Technology, Central Queensland University, Rockhampton, QLD 4701, Australia. email address: (M.A.S): m.sadiq@cqu.edu.au; (P.K.L): p.langat@cqu.edu.au; [ORCID iD: https://orcid.org/0000-0003-0412-4804](https://orcid.org/0000-0003-0412-4804); (A.N): a.neupane@cqu.edu.au

*Correspondence: p.langat@cqu.edu.au

Abstract

Spatial data supply chains (SDSCs) require robust provenance mechanisms to ensure data quality, traceability, and interoperability across geospatial workflows. This study presents a systematic review of semantic web-based approaches to provenance modelling in SDSCs, synthesising evidence from 156 studies published between 2001 and 2025. The review evaluates the use of semantic technologies, including RDF, OWL, SPARQL, and GeoSPARQL, and benchmarks existing provenance models against criteria of granularity, scalability, and standards compliance. The findings reveal fragmented lineage practices, limited feature-level provenance representation, and persistent challenges related to real-time processing, interoperability, and scalability. To address these gaps, the study identifies the need for GeoPROV, a minimal and interoperable semantic framework that extends W3C PROV with spatial semantics while maintaining compatibility with ISO lineage standards and emerging catalogue specifications. GeoPROV can enhance trust in real-world spatial data ecosystems. The review concludes by outlining practical implications for operationalising GeoPROV in SDSCs, identifying research priorities for automated provenance capture and big-data scalability, and highlighting the role of semantic reasoning in improving trust, transparency, and reproducibility in spatial data governance.

Keywords: Spatial data supply chains (SDSCs), Geospatial provenance, spatial data infrastructure (SDI), Geospatial standards, GeoPROV framework

1. Introduction

Spatial data supply chains (SDSCs) are distributed, multi-actor ecosystems. They span the acquisition, transformation, analysis, and dissemination of geospatial data products and services. (1, 2). Unlike traditional supply chains, SDSCs face unique challenges. Coordinate system transformations, spatial resolution changes, multi-scale representations, and the inherent complexity of maintaining lineage across heterogeneous geospatial workflows are some of the challenging tasks (3-7). The provenance of spatial data documenting its data origin, processing history, and transformation steps (8) has become critically important. This is because it ensures information quality, traceability, and trust in decision-making processes (9) that increasingly rely on geospatial intelligence (10).

The evolution of spatial data infrastructures (SDIs) into service-oriented, cloud-native, and platform-based architectures has exponentially increased the complexity of provenance tracking (11). Contemporary spatial data ecosystems involve multiple stakeholders, technologies, and governance frameworks, creating intricate webs of data dependencies that traditional lineage models struggle to capture comprehensively. This complexity is further amplified by the integration of emerging technologies, including Internet of Things (IoT) sensor networks, artificial intelligence (AI) and machine learning (ML) pipelines, and blockchain-based mechanisms for auditability and traceability (12-14). Although provenance has been extensively studied in scientific workflows, sensor networks, and data infrastructures, the integration of Semantic Web technologies with geospatial provenance remains inconsistent. Existing approaches typically combine W3C PROV with geospatial standards such as ISO 19115 lineage elements or GeoSPARQL; however, these integrations are often ad hoc, domain-specific, or insufficiently formalised. As a result, provenance representations are difficult to validate, reuse, or reason over across organisational and disciplinary boundaries, limiting their ability to support automated quality assessment, interoperability, and trust at scale.

This study presents a systematic review of Semantic Web-based approaches to provenance modelling in SDSCs, synthesising evidence literature publications published between 2001 and 2025. The review examines how Semantic Web technologies, including RDF, OWL, SPARQL, GeoSPARQL, and shapes constraint language (SHACL), are used to represent, validate, and query spatial provenance, and evaluates existing models against criteria of granularity, scalability, and standards alignment. Beyond synthesis, the paper identifies recurring ontology design patterns, interoperability limitations, and reasoning constraints that restrict the operational use of provenance in real-world SDSCs.

Based on this analysis, we introduce GeoPROV, a lightweight semantic profile that extends PROV-O with explicit spatial semantics while maintaining compatibility with established Semantic Web and geospatial standards. GeoPROV is positioned not as a replacement for existing models, but as a minimal, interoperable layer that formalises spatial provenance concepts for linked-data publication, constraint validation, and cross-domain reuse. This work articulates design principles and research priorities for provenance-centric spatial knowledge graphs. It advances the role of semantic web technologies in enabling trustworthy, explainable, and interoperable spatial data ecosystems.

The remainder of this paper is organized as follows: Section 2 describes the review methodology and analysis approach, including the PRISMA protocol and coding schema. Section 3 synthesizes key concepts, standards, semantic web technologies, and existing spatial provenance models, highlighting their relevance to Spatial Data Supply Chains (SDSCs). Section 4 examines integration patterns, querying and reasoning capabilities, and interoperability challenges in current frameworks. Section 5 discusses research gaps and limitations, followed by future research priorities. Finally, Section 6

concludes with practical implications of the proposed GeoPROV framework and recommendations for advancing provenance-centric SDSCs.

2. Method and analysis

2.1 Review Method and Protocol

This study employed a systematic literature review guided by PRISMA to ensure rigor and transparency. The protocol targeted research on provenance modeling in SDSCs using semantic web technologies published between 2001 and 2025. The process followed six phases. These included defining research questions and criteria, database searching, screening, full-text assessment, data extraction and synthesis, and narrative analysis to identify patterns and gaps. Figure 1 illustrates the PRISMA flow, including database retrieval (e.g., Scopus, Web of Science, IEEE Xplore), duplicate removal, screening, and final inclusion.

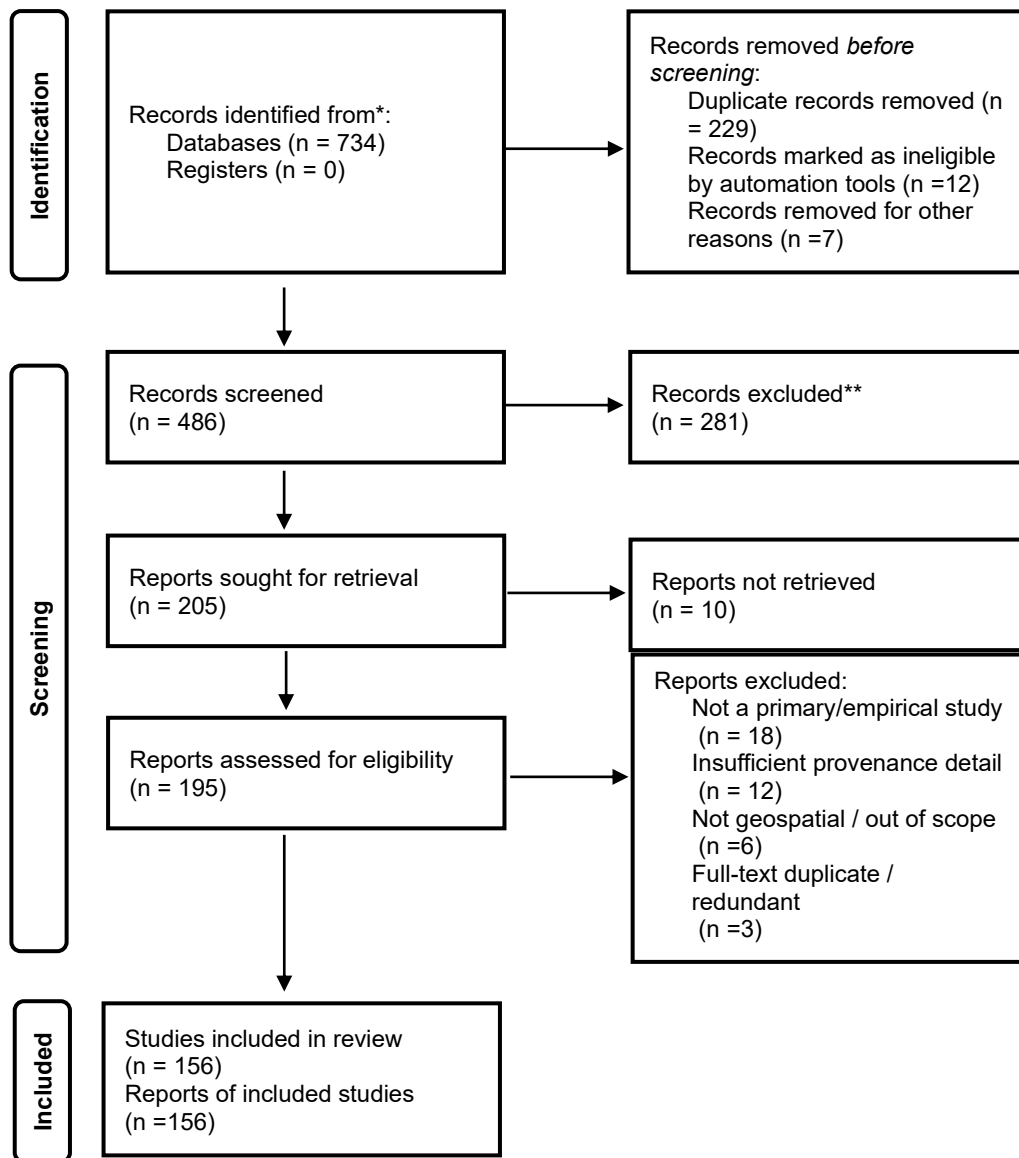


Fig. 1. PRISMA flow diagram illustrating the systematic review process for semantic web-based provenance modelling in spatial data supply chains (SDSCs).

2.2 Analysis approach

The study applied a structured coding framework to classify and interpret literature across three dimensions: domains, capture modes, and representation approaches. A detailed codebook with definitions, examples, and inclusion criteria ensured consistent tagging during screening and extraction. Domains covered earth observation, cadastral systems, hydrology, LADM, BIM, pipelines, smart cities, environmental monitoring, transportation, and disaster response. Capture modes included native tool outputs, orchestrator-level interception, log-to-graph adapters, manual annotation, automated metadata extraction, and service-oriented capture. Representation approaches spanned W3C PROV models, ISO lineage standards, custom ontologies, GeoSPARQL, DCAT/Geodata alignment, SOSA/SSN models, and hybrid solutions. This multi-label coding enabled nuanced insights into provenance modelling, semantic integration, and interoperability challenges in spatial data workflows.

The evolution diagram (Table 1) was derived from systematic review coding. Each study was mapped by year and research focus—foundational provenance, workflow integration, Linked Data ontologies, quality governance, or industry-scale automation. These were grouped into four periods (2007–2011, 2012–2014, 2015–2017, 2018–2025) to reveal thematic shifts. Summaries for each period informed a timeline illustrating progression from foundational models to geospatial specialisation, Linked Data integration, and automated SDSC applications.

Table1: Evolution of Semantic Provenance Research in Spatial Data (2007–2025)

Period	Theme	Summary
2021–2025	Advanced systems, automation & industry applications	Operationalised automated capture (semantic execution engines, SHACL validation), scalable graph storage, and catalogue links (e.g., DCAT/STAC). Demonstrated industry use cases—Earth Observation auditing, climate traceability, smart supply chains—showing improved integrity, interoperability, and adaptability in mission-critical SDSCs.
2015–2020	Domain ontologies and Linked Data for SDSC	Introduced domain-specific extensions (e.g., land administration, biodiversity) and Linked Data publication of provenance. Advanced GeoSPARQL use, catalogue/crawl integration, and workflow chaining. Granularity and end-to-end traceability across SDSCs emerged as central design goals.
2012–2014	Geospatial provenance modelling and workflow integration	Adapted provenance to spatial contexts: feature/attribute level, mappings between ISO lineage and semantic models, and provenance-aware geoprocessing web chains. Emphasis on auditing conflation, CRS/transform steps, and associating lineage with data quality assertions.
2007–2011	Foundational semantic position and architectures	Operationalised semantic foundations for provenance using RDF and early workflow integrations. Defined generic entity-agent patterns and service-aware capture, linking provenance to web resources and catalogues to enable basic interoperability and query.

3. Concepts, standards, semantic web technologies, and existing models.

3.1 SDSC definition and relevance

A SDSC is an end-to-end ecosystem of actors, processes, and technologies that create, transform, and deliver geospatial products and services. Beyond traditional data lineage, SDSCs integrate governance, quality assurance, and trust mechanisms to ensure data fitness for decision-making. They involve complex spatial operations, temporal updates, and thematic enrichment across distributed systems. This process requires machine-actionable provenance that links entities, activities, and agents (e.g., W3C PROV-O) with geospatial semantics (e.g., GeoSPARQL) and metadata standards (e.g., DCAT, ISO 19115/19157). Provenance is critical across domains ranging from earth observation and cadastral workflows to hydrology and environmental monitoring. It also underpins AI and machine learning pipelines by supporting traceability, accountability, and compliance. Aligning semantic provenance with Spatial Data Infrastructure (SDI) and catalogue standards reduces compliance burden, enables defect tracing, and enhances trust in geospatial data ecosystems.

Table 2 summarises major research themes for SDSC provenance. These cover supply chain principles, transparency and trust, and Earth observation platforms. They also include SDSC architectures, interoperability strategies, and lineage standards. Additional themes address feature-level and scalable provenance, machine learning workflows, and quality elements. Governance, compliance, policy frameworks, versioning practices, and smart city data management are also included. To highlight opportunities for advancing semantic provenance modelling, each theme presents representative studies, methodological approaches, comparative contributions, limitations, SDSC relevance, and identified gaps.

The synthesis of key themes highlights important limitations. Foundational supply chain management models, such as SCOR and SCOR digital standard (DS), provide process discipline and traceability. However, they lack spatial semantics and feature-level granularity. These capabilities are essential for spatial data supply chains. Emerging efforts in earth observation (EO) data platforms and SDSC architectures demonstrate progress toward provenance-centric workflows but face scalability, interoperability, and quality integration challenges. Interoperability remains a critical barrier, with RDF-based approaches offering promise yet incurring high integration costs. Lineage standards like International Organisation for Standardisation (ISO) and World Wide Web Consortium Provenance Ontology (W3C PROV) establish canonical structures but fall short in capturing intermediate outputs and dynamic workflows. Advanced themes, such as feature-level provenance, scalable lineage for big spatial data, and ML workflow integration, highlight the need for domain-specific profiles and real-time capture mechanisms. Governance, compliance, and versioning considerations further underscore the importance of harmonizing technical and policy vocabularies. Collectively, these gaps highlight the

need for a comprehensive semantic framework. Such a framework should extend W3C PROV with spatial semantics, support scalable automated lineage capture, and align with standards such as spatial temporal asset catalogue (STAC) to ensure trust, transparency, and interoperability in SDSCs.

Table 2. Summary of key themes, representative studies, methods/standards, contributions, limitations, SDSC relevance, and enhancement opportunities identified.

Theme	Methods standards	Contributions (comparative)	Limitations / critiques	SDSC relevance	Gaps & enhancement opportunities
SCM foundations (15, 16)	SCOR/SCOR-DS	Process and performance discipline; unifies plan, source, make, deliver, and return	Abstract wrt data semantics; limited guidance for feature-level lineage	Frames SDSC processes	Couple SCOR-DS KPIs with lineage KPIs (coverage of provenance, parameter completeness)
Transparency & trust (16, 17)	EPCIS 2.0; DPP	Converts ethics and standards into traceability standards; event-level visibility	Supply-chain oriented; limited geospatial semantics	Provenance expectations baseline	Bridge EPCIS events ↔ PROV-O/ISO lineage (crosswalk profile)
EO data platforms (17-19)	Open-source, NCI, ARD	Makes provenance central to EO data cubes; identifies schema evolution pain	Relational index performance; provenance scale issues	SDSC case study	Graph native lineage; PID-based feature versioning; partial reprocessing plans Map architecture to OGC APIs, GeoSPARQL, and ISO 19157-1; add user-facing fitness-for-use templates
SDSC architecture (20-22)	SDSC samples: user perspective	Finds capability gaps, proposes Common Reference Architecture, and exposes metadata insufficiency	Lacks concrete standard mappings	Direct	Automate schema/project ion harmonisation; publish deltas with PROV.
Interoperability (23-27)	SDI, Linked Data	Documents heterogeneity barriers; shows RDF/SPARQL linking and versioning	High integration cost; patchy quality/updates	Central	

Theme	Methods standards	Contributions (comparative)	Limitations / critiques	SDSC relevance	Gaps & enhancement opportunities
Lineage standards (28, 29)	ISO lineage; quality; PROV	Canonical lineage and quality grammar; parameter capture via LE_ProcessParameter	Intermediate outputs and result validation are still weak.	Core	Define a domain-specific “Geo-PROV profile” for feature/attribute lineage and validation artefacts Add streaming/real-time parameters; standardize step granularity; provenance for conflation at scale
Feature/attribute provenance (30-33)	ISO + WPS/OGC API—Processes + PROV	Demonstrates feature/attribute-level capture and publishing	Workflow dependencies in non-sequential flows	High	A capture profile for spatial ops (operators, tolerances, CRS, indexes) Align ML provenance with spatial features; record feature engineering lineage.
Scalable lineage (34-38)	Sedona; Spark—Atlas	Spatial big-data operators; production lineage capture across pipelines	No domain capture profile; limited quality semantics	High	Bind every metric to explicit lineage step IDs and parameter bundles. Ontology bridges (policy ↔ process); purpose limitation recorded in lineage
ML & workflow provenance (39-41)	Script analysis (Vasa); CWLProv	Extracts data-model relations; interoperable workflow provenance	Code-centric; not feature-centric	Moderate	Mandatory provenance disclosures in DSAs; certify lineage APIs
Quality elements (42-45)	Accuracy/consistency metrics	Operationalises tests and rules for fitness for fitness-for-purpose	Often siloed from lineage	Central	
Governance & compliance (46-48)	Lineage for GRC; GDPR	Demonstrates lineage as a compliance enabler	Cross-walking legal and technical vocabularies is hard	High	
Australian policy (49-52)	Data sharing; certification	Transparency registers; interoperability assurance	Implementation variability	Direct	

Theme	Methods standards	Contributions (comparative)	Limitations / critiques	SDSC relevance	Gaps & enhancement opportunities
Versioning (53-55)	PID/version policies	Multi-level chains; partial/full reprocessing strategies	Tooling fragmentation	High	PID at feature-level; diff/patch lineage; change-impact dashboards
Smart cities (56)	Digital twins; CRS	Machine-readable provenance and CRS/datum history	Cross-domain semantic alignment	High	Auto-capture CRS/datum per step; legal metadata for cadastre

3.2 Supply chain management and SDSC

Traditional supply chain management (SCM) (57) concepts provide a useful foundation for understanding spatial data flows but require significant adaptation to address the unique characteristics of geospatial information (58). Unlike physical goods, spatial data can be consumed simultaneously by multiple users, transformed without depleting the original resource, and combined in complex ways to create new information products (59-65). Consequently, data supply chains have evolved from simple linear models to complex, interconnected networks that reflect the dynamics of modern data ecosystems (66, 67).

SDSCs exhibit distinctive characteristics that go beyond traditional data lineage models. These include spatial transformations (e.g., reprojection, resampling, generalisation), temporal evolution (e.g., versioning, updates, time-series aggregation), and semantic enrichment (61, 68-70) (e.g., classification, feature extraction, spatial reasoning). Governance within SDSCs involves multiple stakeholders with varying authority and responsibilities, creating intricate trust relationships that must be captured and validated through robust provenance mechanisms (29, 33).

SDSC represents the end-to-end pathway through which geospatial information is created, processed, validated, shared, and used across diverse systems and organisations. Similar to traditional supply chains, SDSCs involve multiple actors, data collectors, processors, analysts, custodians, publishers, and end-users, each introducing value, uncertainty, or risk to the evolving dataset. Their importance has grown as spatial information becomes critical for decision-making in domains such as environmental monitoring, transport planning, mining, smart cities, and emergency management. A well-governed SDSC ensures that lineage, quality, and trustworthiness are traceable at every stage, enabling users to assess fitness for purpose. However, findings summarised in Table 3 indicate that while semantic provenance models have advanced, most fail to fully capture the multi-stage, multi-actor nature of

SDSCs. This gap underscores the need for a provenance framework aligned with real operational dynamics, where understanding how data travels is as critical as understanding the data itself.

Table 3. Data quality elements, practical metrics, and recommended Storage Locations

Quality element	Typical metrics/diagnostics	Capture point	Store in...
Positional accuracy	RMSE vs. reference, offset stats, checkpoints	Transform/QA step	ISO 19157 quality section; PROV note on reference used; QA tables
Attribute (thematic) accuracy	Confusion matrix, precision/recall	Transform/QA step	ISO 19157 measures; PROV entity for the validation dataset.
Logical consistency	Topology errors, domain checks	Ingest/Transform	ISO 19157 logical checks; graph constraints
Completeness	Missing features/attributes, coverage	Capture/Ingest	ISO 19157 completeness; cube/tiling coverage metadata
Lineage	Sources, steps, parameters, agents, dates	All steps	ISO 19115-2 lineage; PROV-O graph

3.3 Provenance Concepts and Standards

Provenance in spatial data encompasses both retrospective provenance (documenting what happened) and prospective provenance (8, 71) (recording what should happen) across the entire data lifecycle. It has evolved from simple audit trails to semantic models capable of supporting complex queries about data quality, trustworthiness, and fitness for purpose. Provenance elevates data from a passive input to a strategic asset (26, 72-74). In domains such as autonomous vehicles, underground utilities, and geofencing, decision quality depends on traceable, machine-readable spatial context (75-77). Data lineage records the data's origin, how it was transformed, and where it flows next (5, 78).

Versioning is critical (54) because it shows that sequential histories of evolving data products underpin reproducibility. In Digital Earth Australia (DEA), interdependencies and algorithmic updates require feature-level versioning with persistent identifiers (79, 80). DEA experience emphasises distinguishing between partial and full reprocessing and defining ingest properties coordinate reference system (CRS), resolution, chunking, compression, and CF conventions to stabilise downstream analyses (81).

The W3C PROV standard dominates provenance representation, offering an entity-activity-agent model (7, 33, 82-85). However, applying PROV to spatial data requires geospatial semantics and CRS integration, which are absent from the core model. The focus should increasingly be on combining PROV with domain vocabularies and spatial standards.

Semantic web based SDSC provenance modelling is vital as data volumes surge (86-89). From early sensor web efforts (25, 59, 90-93) to semantic ontologies for interoperability (85, 87, 94-97) , progress

continues. Yet challenges persist in granular capture, dynamic workflows, and semantic integration (98). The absence of standardised, semantically rich models hampers reproducibility, error tracing, and trust in spatial infrastructures.

A framework linking SDSCs, provenance metadata, and semantic web technologies, emphasising ontology-driven representation and querying (94, 99, 100) is crucial. Provenance is essential for transparent, interoperable workflows and contextual quality assessment (101, 102). ISO 19157-1:2023 defines core quality elements, positional accuracy, attribute accuracy, logical consistency, completeness, and lineage each requiring systematic capture and storage. Table 3 summarises practical metrics and storage strategies. These include positional accuracy using RMSE and checkpoints recorded under ISO 19157 with PROV annotations, and attribute accuracy assessed through confusion matrices linked to validation datasets. It also covers logical consistency via ISO rule checks, completeness from coverage metadata at acquisition, and end-to-end lineage captured using ISO 19115-2 structures and PROV-O graphs. Studies demonstrate semantic integration of PROV-O with ISO lineage for feature-level capture and SPARQL-based auditing (33, 103, 104), workflow-centric graph models (33, 90, 105, 106), and scalable big-data lineage solutions (15, 107-110). Despite progress, gaps remain in proactive capture, intermediate result validation, and domain-specific profiles, underscoring the need for standardised, semantically rich provenance frameworks.

3.4 Evolution of semantic provenance Research in spatial data

Research on provenance modelling for spatial data supply chains using semantic web technologies has evolved from foundational frameworks to advanced, domain-specific applications. Earlier studies focused on developing semantic provenance models (82, 90, 99, 111), architectures, and standards for integration to support traceability and data quality assessment (47, 112). Later work shifted towards geospatial workflows, feature-level granularity, and interoperability challenges, leveraging ontologies, linked data, and semantic execution engines (101, 113, 114). Recent developments incorporate big data integration, automated reasoning, and practical implementations across specific domains, underscoring the role of provenance in trust (115), interoperability (85, 101), and data governance (96, 110, 116).

The reviewed literature consistently highlights the importance of semantic web technologies for enhancing provenance modelling in spatial data supply chains, particularly through extensions to standards such as W3C PROV (36, 83-85, 95, 102, 111, 117-119) and ISO geospatial metadata (29, 40, 61, 120). There is broad consensus on the need to capture provenance at multiple levels of granularity, including feature and attribute levels, to ensure data quality and trustworthiness. However, differences emerge regarding the degree of integration with automated geospatial workflows and the sophistication of querying and reasoning capabilities. These variations reflect differences in domain focus, technical scope, and implementation maturity. Interoperability challenges and inconsistencies in standard

compliance also persist, often due to heterogeneous application contexts and evolving semantic frameworks.

3.5 Semantic Web Technologies (RDF, OWL, SPARQL)

The semantic web technology stack provides the foundational technologies for representing, storing, and querying provenance information in machine-readable formats (83, 121-124). RDF serves as the basic data model for expressing provenance relationships, while the OWL enables the definition of domain-specific vocabularies and reasoning rules (108, 125-129). SPARQL Protocol and the (SPARQL provides the query interface for accessing and analysing provenance information across distributed systems(40, 61, 130).

The application of semantic web technologies to spatial data provenance presents both opportunities and challenges. On one hand, the flexibility of RDF enables the representation of complex spatial relationships and transformations that are difficult to capture in traditional relational models. On the other hand, the verbosity of RDF can lead to performance challenges when dealing with large-scale spatial datasets, requiring careful optimization of storage and query strategies. Table 4 shows that most provenance studies rely on the classic Semantic Web stack: RDF (88.8%) for representation, SPARQL (80.9%) for querying, and OWL (71.9%) for ontologies and reasoning. Specialised technologies appear less frequently but serve specific roles: GeoSPARQL (33.7%) for spatial queries, SHACL (20.2%) for validation, and JSON-LD (28.1%) for web-friendly RDF serialisation. Overall, RDF, SPARQL, and OWL form the core, while GeoSPARQL, SHACL, and JSON-LD address spatial, validation, and integration needs.

Table 4: Semantic Technology Adoption Patterns

Technology	Studies (n)	Percentage	Primary Use Cases
RDF	79	88.8%	Basic provenance representation
SPARQL	72	80.9%	Provenance querying and discovery
OWL	64	71.9%	Ontology definition and reasoning
GeoSPARQL	30	33.7%	Spatial queries over provenance
SHACL	18	20.2%	Provenance validation and constraints
JSON-LD	25	28.1%	Web-friendly RDF serialization

3.6 Existing spatial provenance models

Spatial provenance modelling is anchored in three key standards: W3C PROV, ISO lineage and quality standards, and GeoSPARQL. The W3C PROV family provides a technology-neutral model for representing provenance on the web, built around core concepts—entity, activity, and agent, and relations such as `prov:used`, `prov:wasGeneratedBy`, and `prov:wasDerivedFrom` (33, 131). PROV-O, the OWL/RDF ontology, forms the backbone for semantic provenance and is extended by GeoPROV to incorporate geospatial entities, operations, and parameters for SDSCs. ISO 19115 and ISO 19157 complement PROV by defining metadata and data quality standards. ISO 19115 specifies lineage elements (`LI_Lineage`, `LI_ProcessStep`, `LI_Source`) for catalogue-level documentation, while ISO 19157 formalizes quality measures such as completeness, positional accuracy, and logical consistency (ISO, 2013, 2014). These standards serve as the catalogue-facing layer, while PROV-O and GeoSPARQL enable machine-actionable provenance graphs. GeoSPARQL extends RDF/SPARQL with spatial vocabularies and topological functions (132), allowing queries that combine provenance paths with spatial predicates. Integrating these standards addresses heterogeneity in schemas and CRSs (Figure 2) but remains challenged by tooling maturity and feature-level granularity at scale (29, 33, 103, 104, 133).

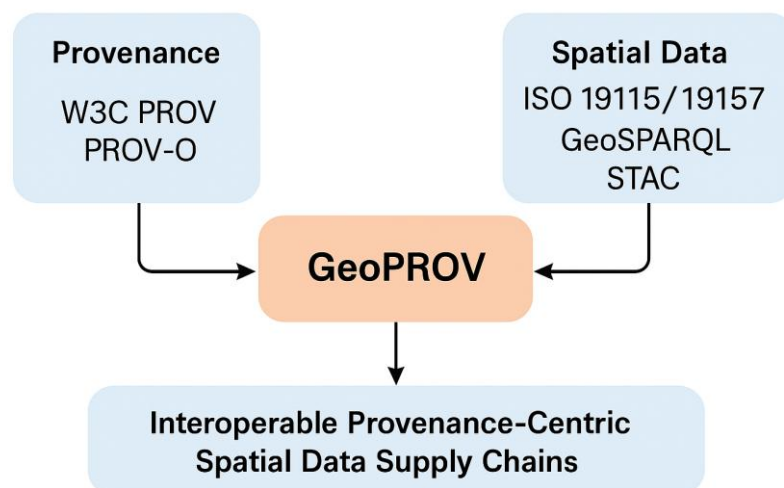


Figure 2. GeoPROV integration conceptual architecture. GeoPROV sits between generic provenance standards (W3C PROV/PROV-O) and geospatial standards (ISO 19115/19157, GeoSPARQL, STAC), providing a minimal, interoperable semantic layer that enables provenance-centric, standards-compliant spatial data supply chains (SDSCs).

Table 5 illustrates the distribution of provenance models and frameworks adopted in spatial data supply chain research, as summarised in Table 4. It compares five categories: W3C PROV/PROV-O, ISO 19115/19157 lineage standards, custom domain ontologies (e.g., GeoPROV, BioProv), the open provenance model (OPM), and hybrid approaches combining multiple standards. The visualisation

highlights W3C PROV as the dominant model, used in 67.4% of studies (n = 60), followed by ISO lineage standards at 44.9% (n = 40). Custom ontologies account for 22.5% (n = 20), while OPM and hybrid approaches represent 16.9% (n = 15) and 13.5% (n = 12), respectively. This figure emphasises the trend toward semantic interoperability, with PROV-based models forming the backbone of most implementations and ISO standards providing complementary, catalogue- lineage and quality structures.

Table 5 Adoption of Provenance Models and Frameworks in SDSC

Model/Framework	Studies (n)	Percentage	Key Characteristics
W3C PROV/PROV-O	60	67.4%	Generic provenance vocabulary, extensible
ISO 19115/19157 Lineage	40	44.9%	Geospatial-specific metadata standards
Custom Domain Ontologies	20	22.5%	Specialised vocabularies (GeoPROV, BioProv)
Open Provenance Model (OPM)	15	16.9%	Legacy provenance model, mapped to PROV
Hybrid Approaches	12	13.5%	Combination of multiple standards

4. Integration, Querying, and Interoperability

4.1 Integration of Semantic Provenance Models in SDSCs

Semantic provenance models are increasingly integrated into SDSCs to support data discovery, traceability, and validation across distributed geospatial workflows (Table 6). In most implementations, provenance is embedded as supplementary metadata alongside datasets or services, rather than treated as a first-class, queryable knowledge graph. Provenance capture is commonly implemented at discrete stages of the data lifecycle, such as ingestion, processing, or publication, using workflow engines, catalogues, or custom middleware components.

While several studies demonstrate successful integration of PROV-based representations within geospatial processing pipelines, these integrations are often tightly coupled to specific platforms or application domains. As a result, provenance graphs are rarely reusable across SDSCs or interoperable between systems. From a Semantic Web perspective, this limits the ability to link provenance entities

across organisational boundaries and to exploit global identifiers, shared vocabularies, and reasoning services. Moreover, the lack of explicit spatial semantics constrains the representation of feature-level lineage and spatial transformations, reducing the explanatory power of provenance graphs in operational settings.

Table 6: Integration Architectural Patterns

Pattern	Studies (n)	Advantages	Limitations
Embedded Capture	32	High completeness, real-time	Tool modification required
Service-Oriented	28	Scalable, standardised	May miss minute details
Post-hoc Reconstruction	21	Non-invasive, retroactive	Incomplete, error-prone
Hybrid Approaches	8	Balanced trade-offs	Complex implementation

4.2 Querying and Reasoning over Spatial Provenance

Querying provenance information is central to its value in SDSCs, enabling users and systems to trace data origins, inspect processing histories, and assess fitness for purpose. Most reviewed studies rely on SPARQL queries over RDF representations of PROV-based graphs, occasionally augmented with GeoSPARQL for spatial filtering and topological relations. In practice, querying is predominantly limited to dataset- or process-level relationships, with limited support for feature-level lineage or parameter-specific interrogation (Table 7). Reasoning over provenance graphs is less commonly implemented and typically restricted to lightweight inferencing, such as subclass reasoning or property inheritance. Although OWL-based reasoning has been proposed to infer implicit dependencies or validate workflow consistency, scalability concerns and the complexity of spatial reasoning often limit adoption. This reflects a broader tension between expressive provenance models and computational tractability, particularly when provenance graphs grow large or are generated continuously. Consequently, many systems prioritise query performance over semantic richness, resulting in underutilisation of reasoning capabilities for trust assessment and automated validation.

Table 7: Querying and Reasoning Capabilities

Capability	Studies (n)	Percentage	Complexity Level
Basic Lineage Traversal	84	94.4%	Simple
Temporal Provenance Queries	37	41.6%	Moderate
Spatial-Provenance Queries	30	33.7%	Moderate
Trust and Quality Assessment	28	31.5%	Complex
Automated Reasoning	34	38.2%	Complex
Cross-dataset Integration	25	28.1%	Complex

About one-third of the reviewed studies (33%) support provenance querying via SPARQL endpoints, semantic web services, and linked data catalogues, enabling the retrieval of detailed metadata. Advanced capabilities include rule-based reasoning for quality inference, formal verification, and explanation generation, complemented by interactive visualisation tools and semantic dashboards for user interpretation. Some frameworks allow complex spatial, temporal, and thematic queries over provenance graphs to support reproducibility and error tracing, though query performance remains a major challenge for large-scale implementations.

4.3 Interoperability across Standards and Platforms

Most studies adopt established standards (Table 8), such as W3C PROV, ISO 19115/19115-2 lineage, OGC specifications, and EPCIS, to ensure interoperability in supply chains (30, 98, 104, 109, 134). Several works propose mappings between standards (e.g., ISO to OPM) to address heterogeneity, while semantic web principles and linked data are used to enhance cross-domain integration and sharing (104, 135). Persistent challenges include standardising provenance granularity and embedding capture seamlessly in diverse geospatial workflows (4, 98, 109). Overall, the analysis reveals a tension between semantic expressiveness and strict standards compliance, often requiring custom extensions or mappings to reconcile generic models with domain-specific needs.

Interoperability remains one of the most persistent challenges in semantic provenance modelling for SDSCs. Existing approaches frequently combine W3C PROV with ISO 19115 lineage elements, GeoSPARQL vocabularies, and domain-specific ontologies. However, these combinations are often implemented through informal mappings or hard-coded transformations rather than explicit, machine-interpretable alignments. As a result, provenance representations remain fragmented, hindering cross-platform reuse and federation.

From a semantic web standpoint, interoperability is further constrained by inconsistent use of identifiers, limited adoption of shared ontology design patterns, and the absence of constraint validation mechanisms. Few studies employ SHACL or similar technologies to enforce structural or semantic consistency in provenance graphs. This gap is particularly problematic for SDSCs operating across organisational or jurisdictional boundaries, where provenance must be exchanged, interpreted, and trusted by heterogeneous stakeholders. Without formalised semantic profiles, provenance models risk becoming bespoke artefacts rather than interoperable components of a broader linked-data ecosystem.

Table 8: Standards Compliance and Interoperability

Standard/Specification	Studies (n)	Compliance Level	Integration Challenges
OGC Web Services	45	High	Legacy system integration
DCAT/GeoDCAT-AP	30	Moderate	Semantic mapping complexity
STAC	11	Emerging	Limited provenance extensions
ISO 19100 Series	52	High	Semantic gap with PROV
Dublin Core	28	Moderate	Limited spatial semantics

4.4 Implications for Scalable and Trustworthy SDSCs

Taken together, the reviewed literature indicates that current semantic provenance implementations fall short of supporting scalable, interoperable, and trustworthy SDSCs. Provenance is often captured retrospectively, queried superficially, and integrated in ways that limit reuse beyond local contexts. These limitations are not inherent to Semantic Web technologies themselves, but rather reflect the absence of lightweight, standardised semantic profiles tailored to spatial data workflows.

This observation motivates the need for provenance models that balance semantic expressiveness with practical deployability. Such models should support feature-level spatial lineage, align explicitly with existing geospatial standards, and remain compatible with scalable RDF storage and querying infrastructures. In this context, the following section introduces GeoPROV as a minimal semantic

profile designed to address these challenges by formalising spatial provenance concepts while preserving interoperability with established Semantic Web standards.

5. Research gaps and limitations of the review.

The literature on SDSC provenance using semantic web technologies shows a strong reliance on standards such as W3C PROV and ISO lineage models, complemented by semantic approaches that enhance interoperability and granularity. Key strengths include ontology-driven provenance capture, advanced querying, and practical implementations that enhance trust and transparency. However, challenges remain in achieving fine-grained provenance, standardisation across heterogeneous systems, and scalability for big data (11, 136, 137). While semantic execution engines and linked data offer promise for automation, adoption is limited by complexity and domain-specific constraints. Overall, the trend points toward more expressive, interoperable, and user-centric frameworks, with gaps in completeness, usability, and cross-domain applicability still to be addressed. Emerging themes address challenges related to provenance granularity, quality assurance, and automation of provenance capture, reflecting an evolving landscape responsive to both theoretical modelling and practical implementation needs (9, 138-140). Table 9 provides a comparative view of methodological trade-offs and implementation gaps identified in the literature.

Table 9: Strengths and weaknesses in provenance automation and big data scalability

Aspect	Strengths	Weaknesses
Provenance modelling frameworks	W3C PROV is widely extended with domain vocabularies (GeoSPARQL, DC/DCAT), yielding expressive, interoperable models for SDSC contexts; land-admin examples show end-to-end chains(141-143)	Feature/attribute-level detail and update/series handling are inconsistent; approaches fragment by domain, complicating reuse. (98, 104, 109)
Integration with geospatial workflows and services	Automatic chaining and provenance-aware GIService/WPS workflows demonstrate capture “where work happens”; linked data publication improves transparency (3, 6, 7).	Heterogeneous platforms and divergent metadata reduce plug-and-play; capture points remain incomplete in many SOA stacks (59).
Granularity and expressiveness	Several works type workflows (tools/algorithms/functions) and attach QA to steps, supporting error tracing and reuse. (3, 39, 104, 144)	Dataset-level provenance dominates; fine-grained (feature/attribute) patterns are rare and burdensome to model/query at scale. (98, 104)

Aspect	Strengths	Weaknesses
Querying, reasoning and visualisation	SPARQL/OWL reasoning enables derivation tracing, quality inference and re-orchestration; catalogue registration and dashboards aid fitness-for-use(61, 145).	Large provenance graphs challenge latency and usability; explanation/UX for non-experts and formal verification remain under-developed (140, 146-148)
Interoperability and standardisation	Alignments among ISO 19115/-2 lineage, OGC specs and PROV exist; mappings (e.g., ISO→OPM/PROV) reduce silos; “common models” encourage continuity (85, 96, 101, 102, 116, 125, 149-152)	Persistent heterogeneity and vendor-specific extensions hinder cross-domain exchange; no universally accepted geo-provenance profile yet. (33, 38, 98, 127)
Automation and semantic execution engines	Runtime capture via semantic engines (61, 92, 153), agents and model-driven harvesting improve completeness and reduces manual effort.	High integration effort in complex, distributed SDSCs; customisation and domain know-how needed; streaming/real-time scenarios stress pipelines (90, 96, 120).
Big-data integration and scalability	Ontology-based integration and linked-data lifecycle tooling show potential for scalable enrichment with provenance control (40, 73, 96).	Reasoning/storage bottlenecks and distributed consistency limit practical deployment at 10^7 – 10^9 -triple scales (154-156).

The practical implications of GeoPROV and Semantic Web technologies extend beyond conceptual modelling to operational use in real-world spatial data supply chains (SDSCs). GeoPROV can be deployed by embedding provenance capture within geoprocessing and analytical environments, enabling automated and standardised recording of lineage during spatial transformations, reprojections, and AI-driven workflows. This supports organisational and regulatory governance, facilitates auditability and defect tracing, and strengthens trust in distributed data ecosystems. Semantic reasoning and constraint validation further enable dynamic assessment of data quality and fitness-for-purpose, allowing decision-makers to query provenance-aware knowledge graphs in near real time. Integration with linked data catalogues and emerging standards such as STAC enhances interoperability across cloud-native infrastructures, while ontology-driven models improve transparency, explainability, and reproducibility in cross-domain workflows.

This review has limitations. The search was restricted to English-language publications, potentially introducing bias, and may not fully capture grey literature or proprietary implementations. Moreover,

the synthesis primarily reflects conceptual contributions, as large-scale empirical deployments remain limited, and the findings represent a snapshot within a rapidly evolving standards landscape.

Future work should focus on automating provenance capture through semantic execution engines to reduce manual effort and enable real-time lineage recording. Achieving feature-level granularity, including CRS and transformation parameters, is critical for dynamic SDSC workflows. Scalability remains a major challenge, requiring distributed graph storage and optimized SPARQL querying for big spatial data. Interoperability must be strengthened by developing a formal GeoPROV profile aligned with W3C PROV, ISO standards, and emerging specifications like STAC. Additionally, integrating semantic reasoning for quality assessment, embedding provenance in AI-driven and cloud-native architectures, and improving visualization tools for non-expert users will enhance usability. Finally, aligning provenance frameworks with governance and compliance requirements will ensure transparency and trust in spatial data ecosystems.

6. Conclusion

Semantic Web technologies offer a strong foundation for modelling provenance in spatial data supply chains (SDSCs), yet their operational uptake remains limited. This systematic review of 156 studies, drawn from an initial corpus of 734 publications, shows that existing approaches—predominantly based on W3C PROV in combination with ISO geospatial metadata standards—tend to capture provenance at coarse dataset levels. Consequently, critical SDSC characteristics such as feature-level lineage, spatial transformations, intermediate processing steps, and dynamic workflow dependencies are often insufficiently represented, constraining interoperability, automated validation, and trust assessment in distributed geospatial environments.

To address these limitations, this paper introduced GeoPROV, a lightweight semantic profile that extends W3C PROV with explicit spatial semantics while remaining compatible with ISO 19115/19157, GeoSPARQL, and emerging catalogue standards such as STAC. GeoPROV provides a minimal, standards-aligned layer for representing machine-actionable spatial provenance, supporting reproducibility, auditability, and cross-domain reuse without imposing excessive modelling overhead. Future research should prioritise automated, workflow-integrated provenance capture, scalable graph-based storage and querying, and semantic reasoning for quality and fitness-for-purpose assessment. Progress will require coordinated engagement between standards bodies, SDSC operators, and the Semantic Web community to translate conceptual advances into interoperable, production-ready provenance solutions for real-world spatial data ecosystems.

Funding

The authors received no financial support for the research, authorship, or publication of this article.

Declaration of Conflicting Interests

The authors declare that there are no conflicts of interest with respect to the research, authorship, or publication of this article.

CRediT authorship contribution statement

Muhammad Azeem Sadiq: Conceptualisation, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualisation, Project administration, Funding acquisition. Philip K. Langat: Supervision, Writing – review & editing, Arjun Neupane: Supervision, Writing – review & editing.

References

1. Varadharajulu P, Azeem Saqiq M, Yu F, McMeekin DA, West G, Arnold L, et al. Spatial data supply chains. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2015;40:41-5.
2. Sadiq MA, West G, McMeekin DA, Arnold L, Moncrieff S. Provenance ontology model for land administration spatial data supply chains. 11th International Conference on Innovations in Information Technology (IIT), IEEE, Dubai, United Arab Emirates, pp 184-189, doi: 10.1109/INNOVATIONS20157381537; 1-3 Nov. 2015: IEEE; 2015. p. 184-9.
3. Scheider S, Ballatore A. Semantic typing of linked geoprocessing workflows. *International Journal of Digital Earth*. 2018;11(1):113-38.
4. Tullis JA, Lanter DP, Basu A, Cothren JD, Shi X, Limp WF, et al. Geoprocessing, workflows, and provenance. *Remote Sensing Handbook, Volume II: CRC Press*; 2024. p. 303-37.
5. Young SG. 13 Geoprocessing, Workflows. *Remote Sensing Handbook, Volume II: Image Processing, Change Detection, GIS, and Spatial Data Analysis*. 2024:303.
6. Yue P, Guo X, Zhang M, Jiang L, Zhai X. Linked Data and SDI: The case on Web geoprocessing workflows. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;114:245-57.
7. Zhang M, Jiang L, Zhao J, Yue P, Zhang X. Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows. *Computers & Geosciences*. 2020;138:104419.
8. Vancisin T, Clarke L, Orr M, Hinrichs U. Provenance visualization: Tracing people, processes, and practices through a data-driven approach to provenance. *Digital Scholarship in the Humanities*. 2023;38(3):1322-39.
9. Tullis JA, Kar B. Where is the provenance? Ethical replicability and reproducibility in GIScience and its critical applications. *Annals of the American Association of Geographers*. 2021;111(5):1318-28.
10. Mai G, Huang W, Sun J, Song S, Mishra D, Liu N, et al. On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv preprint arXiv:230406798*. 2023.
11. Liang H, Zhang Z, Hu C, Gong Y, Cheng D. A survey on spatio-temporal big data analytics ecosystem: Resource management, processing platform, and applications. *IEEE Transactions on Big Data*. 2023;10(2):174-93.
12. Van Hoang T. Impact of integrated artificial intelligence and internet of things technologies on smart city transformation. *Journal of technical education science*. 2024;19(Special Issue 01):64-73.
13. Rane N. Integrating leading-edge artificial intelligence (AI), internet of things (IOT), and big data technologies for smart and sustainable architecture, engineering and construction (AEC) industry: Challenges and future directions. *Engineering and Construction (AEC) Industry: Challenges and Future Directions (September 24, 2023)*. 2023.
14. Ahmed I, Zhang Y, Jeon G, Lin W, Khosravi MR, Qi L. A blockchain-and artificial intelligence-enabled smart IoT framework for sustainable city. *International Journal of Intelligent Systems*. 2022;37(9):6493-507.
15. Guitton F, Oehmichen A, Bossé É, Guo Y. Honest Computing: achieving demonstrable data lineage and provenance for driving data and process-sensitive policies. *Data & Policy*. 2024;6:e84.
16. Yaman B, Thompson K, Fahey F, Brennan R. LinkedDataOps: quality oriented end-to-end geospatial linked data production governance. *Semantic Web*. 2024;15(2):555-81.

17. AlOmeir O. Provenance in relational databases: usability and applications: University of British Columbia; 2023.
18. Sikos LF, Philp D. Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. *Data Science and Engineering*. 2020;5(3):293-316.
19. Gayo JEL, Prud'Hommeaux E, Boneva I, Kontokostas D. *Validating RDF data*: Morgan & Claypool Publishers; 2017.
20. Ataei P, Litchfield A. The state of big data reference architectures: A systematic literature review. *IEEE Access*. 2022;10:113789-807.
21. Haslhofer B, Klas W. A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys (CSUR)*. 2010;42(2):1-37.
22. Martin P, Remy L, Theodoridou M, Jeffery K, Zhao Z. Mapping heterogeneous research infrastructure metadata into a unified catalogue for use in a generic virtual research environment. *Future Generation Computer Systems*. 2019;101:1-13.
23. Abid A, Lee J, Le Gall F, Song J. Toward mapping an NGSI-LD context model on RDF graph approaches: a comparison study. *Sensors*. 2022;22(13):4798.
24. Merrill E, Corlosquet S, Ciccarese P, Clark T, Das S. Semantic Web repositories for genomics data using the eXframe platform. *J Biomed Semantics*. 2014;5(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G):S3.
25. Chatterjee A, Prinz A, Gerdes M, Martinez S. An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study. *J Med Internet Res*. 2021;23(4):e24656.
26. McGlenn K, Rutherford MA, Gisslander K, Hederman L, Little MA, O'Sullivan D. FAIRVASC: A semantic web approach to rare disease registry integration. *Comput Biol Med*. 2022;145:105313.
27. Martins YC, Ziviani A, Cerqueira ECMO, Cavalcanti MCR, Nicolas MF, de Vasconcelos ATR. PPIntegrator: semantic integrative system for protein-protein interaction and application for host-pathogen datasets. *Bioinform Adv*. 2023;3(1):vbad067.
28. Rahima AA, Amalraj AA, Shanofer G, Tamizharasan S, Ganapathy D, Rathnaraj CA. *Geo Spatial Query System Using Quantum NLP. Real-World Applications of Quantum Computers and Machine Intelligence*: IGI Global Scientific Publishing; 2025. p. 105-20.
29. Closa G, Masó J, Zabala A, Pesquer L, Pons X. A provenance metadata model integrating ISO geospatial lineage and the OGC WPS: Conceptual model and implementation. *Transactions in GIS*. 2019;23(5):1102-24.
30. He L, Yue P, Di L, Zhang M, Hu L. Adding geospatial data provenance into SDI—a service-oriented approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2014;8(2):926-36.
31. Malone J, Brown A, Lister AL, Ison J, Hull D, Parkinson H, et al. The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. *J Biomed Semantics*. 2014;5:25.
32. Hasnain A, Mehmood Q, Sana EZS, Saleem M, Warren C, Jr., Zehra D, et al. BioFed: federated query processing over life sciences linked open data. *J Biomed Semantics*. 2017;8(1):13.
33. Closa G, Masó J, Proß B, Pons X. W3C PROV to describe provenance at the dataset, feature and attribute levels in a distributed environment. *Computers, Environment and Urban Systems*. 2017;64:103-17.
34. Boulos MN, Roudsari AV, Carson ER. Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. *Med Sci Monit*. 2002;8(7):MT124-36.
35. Frey JG. The value of the Semantic Web in the laboratory. *Drug Discov Today*. 2009;14(11-12):552-61.
36. Sahoo SS, Valdez J, Rueschman M. Scientific Reproducibility in Biomedical Research: Provenance Metadata Ontology for Semantic Annotation of Study Description. *AMIA Annu Symp Proc*. 2016;2016:1070-9.
37. Ceravolo P, Azzini A, Angelini M, Catarci T, Cudré-Mauroux P, Damiani E, et al. Big data semantics. *Journal on Data Semantics*. 2018;7(2):65-85.
38. Arnaud E, Laporte MA, Kim S, Aubert C, Leonelli S, Miro B, et al. The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. *Patterns (N Y)*. 2020;1(7):100105.

39. Caron C, Amann B, Constantin C, Giroux P, Santanchè A, editors. Provenance-Based Quality Assessment and Inference in Data-Centric Workflow Executions. On the Move to Meaningful Internet Systems: OTM 2014 Conferences; 2014 2014//; Berlin, Heidelberg: Springer Berlin Heidelberg.
40. Gandon F, Guéret C, Villata S, Breslin J, Faron-Zucker C, Zimmermann A. The Semantic Web: ESWC 2015 Satellite Events : ESWC 2015 Satellite Events, Portorož, Slovenia, May 31 – June 4, 2015, Revised Selected Papers [text]. Cham: Imprint: Springer; 2015.
41. Vogt L, Baum R, Bhatti P, Kohler C, Meid S, Quast B, et al. SOCCOMAS: a FAIR web content management system that uses knowledge graphs and that is based on semantic programming. Database (Oxford). 2019;2019.
42. Krishnaswamy P. Winning with DataOps: Harnessing Efficiency in the Enterprise: Libertatem Media Private Limited; 2023.
43. A. Wentz E, Shimizu M. Measuring spatial data fitness-for-use through multiple criteria decision making. *Annals of the American Association of Geographers*. 2018;108(4):1150-67.
44. Bielecka E. GEOGRAPHICAL DATA SETS FITNESS OF USE EVALUATION. *Geodetski vestnik*. 2015;59(2).
45. Devillers R, Bédard Y, Jeansoulin R, Moulin B. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*. 2007;21(3):261-82.
46. Buckley G. Privacy at the intersection of technology, business and regulation: A case study of the GDPR: UCL (University College London); 2025.
47. Vandenbroucke D, Olijslagers M, Boguslawski R, Borzachiello M, Perego A, Smith R. Architectures and Standards for Spatial Data Infrastructures and Digital Government. KJ-NA-30336-EN-N (online). ISSN: 1831-9424 (online). DOI: <https://doi.org> ...; 2020. Report No.: 9276214046.
48. Meijering L, Osborne T, Hoorn E, Montagner C. How the GDPR can contribute to improving geographical research. *Geoforum*. 2020;117:291-5.
49. Box P. The governance of spatial data infrastructure: A registry based model: University of Melbourne, Department of Infrastructure Engineering Melbourne; 2013.
50. Wallace J, Williamson I, Rajabifard A, Bennett R. Spatial information opportunities for Government. *Journal of spatial science*. 2006;51(1):79-99.
51. Woodgate P, Coppa I, Choy S, Phinn S, Arnold L, Duckham M. The Australian approach to geospatial capabilities; positioning, earth observation, infrastructure and analytics: issues, trends and perspectives. *Geo-spatial information science*. 2017;20(2):109-25.
52. Ali ME, Cheema MA, Hashem T, Ulhaq A, Babar MA. Enabling spatial digital twins: Technologies, challenges, and future research directions. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*. 2024;92(6):761-78.
53. Wilson JP, Butler K, Gao S, Hu Y, Li W, Wright DJ. A five-star guide for achieving replicability and reproducibility when working with GIS software and algorithms. *Annals of the American Association of Geographers*. 2021;111(5):1311-7.
54. Klump J, Wyborn L, Wu M, Martin J, Downs RR, Asmi A. Versioning data is about more than revisions: A conceptual framework and proposed principles. *Data Science Journal*. 2021;20:12-.
55. Legge S, Rumpff L, Garnett ST, Woinarski JC. Loss of terrestrial biodiversity in Australia: magnitude, causation, and response. *Science*. 2023;381(6658):622-31.
56. Tran B-H, Aussenac-Gilles N, Comparot C, Trojahn C. Semantic Integration of Raster Data for Earth Observation on Territorial Units. *ISPRS International Journal of Geo-Information*. 2022;11(2):149.
57. Obade VP, Gaya C. Digital technology dilemma: on unlocking the soil quality index conundrum. *Bioresour Bioprocess*. 2021;8(1):6.
58. Alamri S. The geospatial crowd: emerging trends and challenges in crowdsourced spatial analytics. *ISPRS International Journal of Geo-Information*. 2024;13(6):168.
59. International Conference on Conceptual Modeling, Trujillo J. Advances in Conceptual Modeling – Applications and Challenges : ER 2010 Workshops ACM-L, CMLSA, CMS, DE@ER, FP-UML, SeCoGIS, WISM, Vancouver, BC, Canada, November 1-4, 2010, Proceedings [text]. Berlin, Heidelberg: Imprint: Springer; 2010.
60. Brandt AJ, Seabloom EW. Regional and decadal patterns of native and exotic plant coexistence in California grasslands. *Ecol Appl*. 2011;21(3):704-14.

61. Cudre-Mauroux P. The Semantic Web -- ISWC 2012 : 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I [text]. Berlin, Heidelberg: Imprint: Springer; 2012.
62. Li D, Wang S, Li D. Spatial data mining: Springer; 2015.
63. Edwards TR, Armstrong BJ, Birkett-Rees J, Blackwood AF, Herries AIR, Penzo-Kajewski P, et al. Combining legacy data with new drone and DGPS mapping to identify the provenance of Plio-Pleistocene fossils from Bolt's Farm, Cradle of Humankind (South Africa). *PeerJ*. 2019;7:e6202.
64. Kootker LM, Plomp E, Ammer STM, Hoogland V, Davies GR. Spatial patterns in (87)Sr/(86)Sr ratios in modern human dental enamel and tap water from the Netherlands: Implications for forensic provenancing. *Sci Total Environ*. 2020;729:138992.
65. Gopi K, Mazumder D, Crawford J, Gadd P, Tadros CV, Atanacio A, et al. Developing a MySQL Database for the Provenance of Black Tiger Prawns (*Penaeus monodon*). *Foods*. 2023;12(14).
66. Gelhaar J, Otto B. Challenges in the Emergence of Data Ecosystems. *PACIS*. 2020;175.
67. Legenvre H, Hameri A-P. The emergence of data sharing along complex supply chains. *International Journal of Operations & Production Management*. 2024;44(1):292-7.
68. Shotton D, Portwin K, Klyne G, Miles A. Adventures in semantic publishing: exemplar semantic enhancements of a research article. *PLoS Comput Biol*. 2009;5(4):e1000361.
69. Simeone D, Cursi S, Acierno M. BIM semantic-enrichment for built heritage representation. *Automation in Construction*. 2019;97:122-37.
70. Höltingen L, Zentgraf S, Hagedorn P, König M. Utilizing large language models for semantic enrichment of infrastructure condition data: a comparative study of GPT and Llama models. *AI in Civil Engineering*. 2025;4(1).
71. Herschel M, Diestelkämper R, Ben Lahmar H. A survey on provenance: What for? What form? What from? *The VLDB Journal*. 2017;26(6):881-906.
72. Pan B, Stakhanova N, Ray S. Data Provenance in Security and Privacy. *ACM Computing Surveys*. 2023;55(14s):1-35.
73. Bhattacharjee T, Kiwuwa-Muyingo S, Kanjala C, Maoyi ML, Amadi D, Ochola M, et al. INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools. *Front Digit Health*. 2024;6:1329630.
74. Bukhari TT, Oladimeji O, Etim ED, Ajayi JO. Systematic review of metadata-driven data orchestration in modern analytics engineering. *Gyanshauryam, International Scientific Refereed Research Journal*. 2022;5(4):536-64.
75. Fisher M, Mascardi V, Rozier KY, Schlingloff B-H, Winikoff M, Yorke-Smith N. Towards a framework for certification of reliable autonomous systems. *Autonomous Agents and Multi-Agent Systems*. 2021;35(1):8.
76. Triapitcin I. Knowledge Injection Method for Real-Time Decision Support. 2023.
77. Amadi D, Kiwuwa-Muyingo S, Bhattacharjee T, Taylor A, Kiragga A, Ochola M, et al. Making Metadata Machine-Readable as the First Step to Providing Findable, Accessible, Interoperable, and Reusable Population Health Data: Framework Development and Implementation Study. *Online Journal of Public Health Informatics*. 2024;16(1):e56237.
78. Olaya P, Kennedy D, Llamas R, Valera L, Vargas R, Lofstead J, et al. Building trust in earth science findings through data traceability and results explainability. *IEEE Transactions on Parallel and Distributed Systems*. 2022;34(2):704-17.
79. Zhang C, Bijlard J, Staiger C, Scollen S, van Enckevort D, Hoogstrate Y, et al. Systematically linking tranSMART, Galaxy and EGA for reusing human translational research data. *F1000Res*. 2017;6.
80. Hojati M, Robertson C, Roberts S, Chaudhuri C. GIScience research challenges for realizing discrete global grid systems as a Digital Earth. *Big Earth Data*. 2022;6(3):358-79.
81. Owers CJ, Lucas RM, Clewley D, Tissott B, Chua SM, Hunt G, et al. Operational continental-scale land cover mapping of Australia using the Open Data Cube. *International Journal of Digital Earth*. 2022;15(1):1715-37.
82. Jayapandian CP, Zhao M, Ewing RM, Zhang GQ, Sahoo SS. A semantic proteomics dashboard (SemPoD) for data management in translational research. *BMC Syst Biol*. 2012;6 Suppl 3(Suppl 3):S20.
83. Wang X, Wang Y, Ambite JL, Appaji A, Lander H, Moore SM, et al. Enabling Scientific Reproducibility through FAIR Data Management: An ontology-driven deep learning approach in the NeuroBridge Project. *AMIA Annu Symp Proc*. 2022;2022:1135-44.

84. Plass M, Wittner R, Holub P, Frexia F, Mascia C, Gallo M, et al. Provenance of specimen and data - A prerequisite for AI development in computational pathology. *N Biotechnol.* 2023;78:22-8.
85. Prudhomme T, De Colle G, Liebers A, Sculley A, Xie PK, Cohen S, et al. A semantic approach to mapping the Provenance Ontology to Basic Formal Ontology. *Sci Data.* 2025;12(1):282.
86. Parikh PP, Minning TA, Nguyen V, Lalithsena S, Asiaee AH, Sahoo SS, et al. A semantic problem solving environment for integrative parasite research: identification of intervention targets for *Trypanosoma cruzi*. *PLoS Negl Trop Dis.* 2012;6(1):e1458.
87. Vos RA, Biserkov JV, Balech B, Beard N, Blissett M, Brenninkmeijer C, et al. Enriched biodiversity data as a resource and service. *Biodivers Data J.* 2014(2):e1125.
88. Gierend K, Kruger F, Genehr S, Hartmann F, Siegel F, Waltemath D, et al. Provenance Information for Biomedical Data and Workflows: Scoping Review. *J Med Internet Res.* 2024;26:e51297.
89. Klingner CM, Denker M, Grun S, Hanke M, Oeltze-Jafra S, Ohl FW, et al. Research Data Management and Data Sharing for Reproducible Research-Results of a Community Survey of the German National Research Data Infrastructure Initiative Neuroscience. *eNeuro.* 2023;10(2).
90. Fonseca F, Rodríguez MA, Levashkin S. *GeoSpatial semantics : second international conference, GeoS 2007, Mexico City, Mexico, November 29-30, 2007 : proceedings [text].* New York, New York: Springer; 2007.
91. Basu A, Das S, Horain P, Bhattacharya S, SpringerLink (Online service). *Intelligent human computer interaction : 8th International Conference, IHCI 2016, Pilani, India, December 12-13, 2016, proceedings [text].* Cham: Imprint: Springer; 2017. Available from: <http://dx.doi.org/10.1007/978-3-319-52503-7>.
92. Gonzalez-Gil P, Martinez JA, Skarmeta AF. *Lightweight Data-Security Ontology for IoT. Sensors (Basel).* 2020;20(3).
93. Chen JYC, Fragoneni G, Fang X. *HCI International 2023 - Late Breaking Papers : 25th International Conference on Human-Computer Interaction, HCII 2023, Copenhagen, Denmark, July 23-28, 2023, Proceedings, Part V [text].* Cham, Switzerland: Springer; 2023.
94. Sahoo SS, Nguyen V, Bodenreider O, Parikh P, Minning T, Sheth AP. A unified framework for managing provenance information in translational research. *BMC Bioinformatics.* 2011;12:461.
95. Ciccarese P, Soiland-Reyes S, Belhajjame K, Gray AJ, Goble C, Clark T. PAV ontology: provenance, authoring and versioning. *J Biomed Semantics.* 2013;4(1):37.
96. Ambalavanan R, Snead RS, Marczika J, Towett G, Malioukis A, Mbogori-Kairichi M. Ontologies as the semantic bridge between artificial intelligence and healthcare. *Front Digit Health.* 2025;7:1668385.
97. Lin CL, Huang PC, Woll C, Theato P, Kubel C, Pilz L, et al. Addressing standardization and semantics in an electronic lab notebook for multidisciplinary use: LabIMotion. *J Cheminform.* 2025;17(1):75.
98. Sadiq MA, McMeekin D, Arnold L. Spatial data supply chain provenance modelling for next generation spatial infrastructures using semantic web technologies. *GEOProcessing 2016.* 2016:157.
99. Hastings J, Chepelev L, Willighagen E, Adams N, Steinbeck C, Dumontier M. The chemical information ontology: provenance and disambiguation for chemical data on the biological semantic web. *PLoS One.* 2011;6(10):e25513.
100. Nguyen V, Bodenreider O, Sheth A. Don't Like RDF Reification? Making Statements about Statements Using Singleton Property. *Proc Int World Wide Web Conf.* 2014;2014:759-70.
101. Babb L, Bult C, Carey VJ, Carroll RJ, Hitz BC, Mungall CJ, et al. Improving the FAIRness and Sustainability of the NHGRI Resources Ecosystem. *ArXiv.* 2025.
102. Khan FZ, Soiland-Reyes S, Sinnott RO, Lonie A, Goble C, Crusoe MR. Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv. *Gigascience.* 2019;8(11).
103. Closa G, Masó J, Pesquer L, Pons X, editors. *Auditing Remote Sensing Data Using Geospatial Provenance. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium; 2019 28 July-2 Aug. 2019.*
104. Closa G, Masó J, Julià N, Pons X. Geospatial queries on data collection using a common provenance model. *ISPRS International Journal of Geo-Information.* 2021;10(3):139.

105. Meersman R, Panetto H, Dillon T, Missikoff M, Liu L, Pastor O, et al. On the Move to Meaningful Internet Systems: OTM 2014 Conferences : Confederated International Conferences: CoopIS and ODBASE 2014, Amantea, Italy, October 27-31, 2014. Proceedings [text]. Berlin, Heidelberg: Imprint: Springer; 2014.
106. Elshehaly M, Jianu R, Slingsby A, Andrienko G, Andrienko N, Rhyne TM. Designing for Collaboration: Visualization to Enable Human-LLM Analytical Partnership. *IEEE Comput Graph Appl.* 2025;45(5):107-16.
107. Aravind L, Iyer LM. Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. *Cell Cycle.* 2003;2(4):369-76.
108. Sahoo SS, Bodenreider O, Hitzler P, Sheth A, Thirunarayan K. Provenance Context Entity (PaCE): Scalable Provenance Tracking for Scientific RDF Data. *Sci Stat Database Manag.* 2010;6187:461-70.
109. Di L, Shao Y, Kang L. Implementation of geospatial data provenance in a web service workflow environment with ISO 19115 and ISO 19115-2 lineage model. *IEEE transactions on geoscience and remote sensing.* 2013;51(11):5082-9.
110. Verma R, Shrivastava P, Merla N. Tracing the path: Data lineage and its impact on data governance. *International Journal of Global Innovations and Solutions (IJGIS).* 2024.
111. Deus HF, Prud'hommeaux E, Miller M, Zhao J, Malone J, Adamusiak T, et al. Translating standards into practice - one Semantic Web API for Gene Expression. *J Biomed Inform.* 2012;45(4):782-94.
112. Castano S. Advances in conceptual modeling : ER 2012 workshops: CMS, ECDM-NoCoDA, MoDIC, MORE-BI, RIGiM, SeCoGIS, WISM, Florence, Italy, October 15-18, 2012 : proceedings [text]. Berlin ; New York: Springer,; 2012.
113. Yang Y, Wang S, Chen Y, Wang X, Jiang W, Jin Y, et al. Ontolomics-P: Advancing Proteomics Data Interpretation through GPT-4o Reannotated Topic Ontology and Data-Driven Analysis. *Anal Chem.* 2025;97(19):10299-308.
114. Nisar M, Soman SP, Sreelan S, John L, Pinto SM, Kandasamy RK, et al. ProteoArk: A One-Pot Proteomics Data Analysis and Visualization Tool for Biologists. *J Proteome Res.* 2025;24(3):1008-16.
115. Kale A, Nguyen T, Harris FC, Li C, Zhang J, Ma X. Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence.* 2023;5(1):139-62.
116. Basereh M, Caputo A, Brennan R. Automatic transparency evaluation for open knowledge extraction systems. *J Biomed Semantics.* 2023;14(1):12.
117. Chen C, Huang H, Ross KE, Cowart JE, Arighi CN, Wu CH, et al. Protein ontology on the semantic web for knowledge discovery. *Sci Data.* 2020;7(1):337.
118. Hussain S, Sun H, Sinaci A, Erturkmen GB, Mead C, Gray AJ, et al. A framework for evaluating and utilizing medical terminology mappings. *Stud Health Technol Inform.* 2014;205:594-8.
119. Sahoo SS, Valdez J, Kim M, Rueschman M, Redline S. ProvCaRe: Characterizing scientific reproducibility of biomedical research studies using semantic provenance metadata. *Int J Med Inform.* 2019;121:10-8.
120. Soranno PA, Bissell EG, Cheruvelil KS, Christel ST, Collins SM, Fergus CE, et al. Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. *Gigascience.* 2015;4:28.
121. Li H, Yue P, Wu H, Teng B, Zhao Y, Liu C. A question-answering framework for geospatial data retrieval enhanced by a knowledge graph and large language models. *International Journal of Digital Earth.* 2025;18(1).
122. Liu Y, Wang X, Wang Y, Huang F, Huang Y, Li Y, et al. Representation learning for geospatial data. *Annals of GIS.* 2025:1-27.
123. Wang L, Ambite JL, Appaji A, Bijsterbosch J, Dockes J, Herrick R, et al. NeuroBridge: a prototype platform for discovery of the long-tail neuroimaging data. *Front Neuroinform.* 2023;17:1215261.
124. Yue Z, Willey CD, Hjelmeland AB, Chen JY. BEERE: a web server for biomedical entity expansion, ranking and explorations. *Nucleic Acids Res.* 2019;47(W1):W578-W86.
125. Anguita A, Garcia-Remesal M, de la Iglesia D, Graf N, Maojo V. Toward a view-oriented approach for aligning RDF-based biomedical repositories. *Methods Inf Med.* 2015;54(1):50-5.

126. Bolling C, Weidlich M, Holzhutter HG. SEE: structured representation of scientific evidence in the biomedical domain using Semantic Web techniques. *J Biomed Semantics*. 2014;5(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G):S1.
127. Chalk SJ. SciData: a data model and ontology for semantic representation of scientific data. *J Cheminform*. 2016;8:54.
128. Fasquelle-Lopez J, Louis Raisaro J. An Ontology and Data Converter from RDF to the i2b2 Data Model. *Stud Health Technol Inform*. 2022;294:372-6.
129. Fu G, Batchelor C, Dumontier M, Hastings J, Willighagen E, Bolton E. PubChemRDF: towards the semantic annotation of PubChem compound and substance databases. *J Cheminform*. 2015;7:34.
130. Tong Q, He D, Xia Z, Huang J, Li Y, Xu F, et al. Research on sand body architecture at the intersection of a bidirectional sedimentary system in the Jiyuan area of Ordos Basin. *Sci Rep*. 2023;13(1):1226.
131. Lebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, Corsar D, et al. Prov-o: The prov ontology. 2013.
132. Zhang F, Lu Q, Du Z, Chen X, Cao C. A comprehensive overview of RDF for spatial and spatiotemporal data management. *The Knowledge Engineering Review*. 2021;36:e10.
133. Wiemann S, Bernard L. Spatial data fusion in Spatial Data Infrastructures using Linked Data. *International Journal of Geographical Information Science*. 2015;30(4):613-36.
134. Yue P, Wei Y, Di L, He L, Gong J, Zhang L. Sharing geospatial provenance in a service-oriented environment. *Computers, Environment and Urban Systems*. 2011;35(4):333-43.
135. Curry E, O'Donnell J, Corry E, Hasan S, Keane M, O'Riain S. Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics*. 2013;27(2):206-19.
136. Ruan P, Chen G, Dinh TTA, Lin Q, Ooi BC, Zhang M. Fine-grained, secure and efficient data provenance on blockchain systems. *Proceedings of the VLDB Endowment*. 2019;12(9):975-88.
137. Jiang L, Yue P, Kuhn W, Zhang C, Yu C, Guo X. Advancing interoperability of geospatial data provenance on the web: Gap analysis and strategies. *Computers & Geosciences*. 2018;117:21-31.
138. Pérez B, Rubio J, Sáenz-Adán C. A systematic review of provenance systems. *Knowledge and Information Systems*. 2018;57(3):495-543.
139. Lemieux VL, Group i. Provenance: Past, present and future in interdisciplinary and multidisciplinary perspective. *Building Trust in Information: Perspectives on the Frontiers of Provenance*: Springer; 2016. p. 3-45.
140. Fischer J, Egli L, Groth J, Barrasso C, Ehrmann S, Figgemeier H, et al. Approaches and tools for user-driven provenance and data quality information in spatial data infrastructures. *International Journal of Digital Earth*. 2023;16(1):1510-29.
141. Missier P, Belhajjame K, Cheney J, editors. The W3C PROV family of specifications for modelling provenance metadata. *Proceedings of the 16th international conference on extending database technology*; 2013.
142. Li C, Sugimoto S. Provenance description of metadata vocabularies for the long-term maintenance of metadata. *Journal of Data and Information Science Vol*. 2017;2(2).
143. Garijo D, Poveda-Villalón M. Best practices for implementing fair vocabularies and ontologies on the web. *Applications and practices in ontology design, extraction, and reasoning*: IOS Press; 2020. p. 39-54.
144. Vogt L, Kuhn T, Hoehndorf R. Semantic units: organizing knowledge graphs into semantically meaningful units of representation. *J Biomed Semantics*. 2024;15(1):7.
145. Kühlmann BM. Implementing the OWL Reasoner in the Virtual Knowledge Graph System *Ontop*: The University of Bergen; 2024.
146. Radosevic N, Duckham M, Saiedur Rahaman M, Ho S, Williams K, Hashem T, et al. Spatial data trusts: an emerging governance framework for sharing spatial data. *International Journal of Digital Earth*. 2023;16(1):1607-39.
147. Ding H, Zhai J, Deng D, Ma S, editors. The case for learned provenance graph storage systems. *32nd USENIX Security Symposium (USENIX Security 23)*; 2023.
148. Janowicz K, Hitzler P, Li W, Rehberger D, Schildhauer M, Zhu R, et al. Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. *AI Magazine*. 2022;43(1):30-9.

149. Wood EC, Glen AK, Kvarfordt LG, Womack F, Acevedo L, Yoon TS, et al. RTX-KG2: a system for building a semantically standardized knowledge graph for translational biomedicine. *BMC Bioinformatics*. 2022;23(1):400.
150. The Semantic Web -- ISWC 2012 : 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part II [text]. Berlin, Heidelberg: Imprint: Springer; 2012.
151. Bai J, Mosbach S, Taylor CJ, Karan D, Lee KF, Rihm SD, et al. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun*. 2024;15(1):462.
152. Kersloot MG, Jacobsen A, Groenen KHJ, Dos Santos Vieira B, Kaliyaperumal R, Abu-Hanna A, et al. De-novo FAIRification via an Electronic Data Capture system by automated transformation of filled electronic Case Report Forms into machine-readable data. *J Biomed Inform*. 2021;122:103897.
153. Portokallidis N, Drosatos G, Kaldoudi E. Capturing Provenance, Evolution and Modification of Clinical Protocols via a Heterogeneous, Semantic Social Network. *Stud Health Technol Inform*. 2016;225:592-6.
154. Krechowicz A, Deniziak S, Łukawski G. Highly scalable distributed architecture for NoSQL datastore supporting strong consistency. *IEEE Access*. 2021;9:69027-43.
155. Honar Pajoo H, Rashid MA, Alam F, Demidenko S. IoT Big Data provenance scheme using blockchain on Hadoop ecosystem. *Journal of Big Data*. 2021;8(1):114.
156. Aviv I, Barger A, Kofman A, Weisfeld R. Reference architecture for blockchain-native distributed information system. *IEEE Access*. 2023;11:4838-51.