applied system innovation

MDPI

Article

Dx-Onto: A Core Ontology for a Semantic-Based Framework for Managing Digital Transformation Projects

Sareeya Ben-arlee [1,2] and Chinnapong Angsuchotmetee [1,3,*]

[1] College of Digital Science, Prince of Songkla University, Songkhla 90110, Thailand; sareeya.be@depa.or.th
[2] Digital Economy Promotion Agency (DEPA), Ministry of Digital Economy and Society, Bangkok 10900, Thailand
[3] Division of Computational Science, Faculty of Science, Prince of Songkla University, Songkhla 90110, Thailand
* Correspondence: chinnapong.a@psu.ac.th

Abstract

The rapid growth of digital transformation (Dx) initiatives across sectors has created an urgent need for structured, scalable, and accurate management of project knowledge. Without effective organization, valuable insights from Dx projects remain fragmented, limiting their reuse and hindering informed decision-making. This research addresses the gap by designing, developing, and validating Dx-Onto, a domain-specific core ontology implemented in OWL and purpose-built for representing and managing knowledge about Dx projects. Dx-Onto models entities, relationships, and attributes from diverse project documentation into a unified knowledge graph, enabling semantic search, cross-project analysis, and context-aware retrieval. To assess performance, a two-pronged evaluation strategy was adopted: (1) scalability experiments using synthetic datasets measured query execution times across volumes ranging from 10 to 1000 projects, and (2) a comparative benchmark against the Core Ontology of Organisational Transformation (COOT) was conducted using a heterogeneous real-world corpus of Thai digital transformation documents. The results confirm Dx-Onto's capacity to scale and demonstrate a higher domain fit (85.2% vs. 77.3%) and superior analytical utility—including transformation-phase and strategic-dimension diagnostics that are structurally impossible under a general-purpose baseline. By positioning Dx-Onto as the core semantic layer for a future Hybrid LLM-Ontology framework, this work lays the groundwork for intelligent, scalable, and reliable knowledge management solutions in the digital transformation domain.

Keywords: ontology; digital transformation; knowledge management; semantic search; Large Language Model

Check for updates

Academic Editor: Teen-Hang Meen

Received: 13 April 2026
Revised: 18 June 2026
Accepted: 29 June 2026
Published: 8 July 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Digital transformation (Dx) projects are complex initiatives that aim to improve organizational performance through digital technologies and strategic changes [1]. These initiatives typically involve multiple departments and stakeholders, progressing through three main stages: digitization (converting analog content to digital form), digitalization (applying digital tools to existing processes), and digital transformation (redefining business models) [2]. Along this journey, a diverse array of heterogeneous artifacts is produced, including proposals, performance data spreadsheets, source code, and multimedia recordings. However, because these artifacts are often distributed across silos and stored in inconsistent formats, organizations face challenges in knowledge management that can hinder consolidation and reuse [3,4].

Challenge 1: Data Heterogeneity. The inherently multidimensional nature of Dx projects leads to the generation of highly heterogeneous data—spanning structured tables, unstructured narrative reports, and multimedia files—which poses significant challenges for systematic knowledge retrieval and retrospective analysis [5–7]. In many public sector contexts, project records are stored in formats ranging from PDF master plans to Excel spreadsheets and video recordings, making it difficult to extract valuable lessons learned or identify reusable best practices across departments.

Challenge 2: Knowledge Silos and Fragmentation. Current reports suggest that ineffective management of project information and a lack of shared semantic understanding often hinder the ability to consolidate diverse datasets [8,9]. Project knowledge is typically scattered across departmental repositories without cross-referencing mechanisms, leading to redundant efforts and the repetition of similar technical oversights. This fragmentation poses a particular challenge for practitioners and government agencies who need to review past Dx projects to support planning and policy-making [10].

Challenge 3: Semantic Interoperability Gap. To illustrate the practical impact, consider a motivating scenario where a large enterprise launches an AI-driven customer service platform. The project team seeks to learn from past initiatives, but inconsistent naming—e.g., "Customer 360 Analytics" vs. "Analytics Platform Phase II"—and technical details fragmented across unstructured emails and meeting notes hinder effective retrieval [11]. Such data management challenges are further evident in Figure 1, where structural variations in digital provider information storage between Organization A (checkbox-based) and Organization B (free-text) hinder interoperability and lead to integration difficulties [12].

No.	Company Name	Member ID	Province	Website	Phone	Email	Internal Mgmt	Production Mgmt	Logistics	HR Mgmt	Marketing	Finance	Other Solutions
1	Company A	202000250	Bangkok	www.focus.com	024400020	case@focus.com	*						
2	Company B	202000252	Bangkok	www.bit.com	0904433333	info@bit.com		*					
3	Company C	202000248	Bangkok	www.solution.com	0818812311	contact@sol.com			*				

No.	Member ID	Company Name	Province	Website	Phone	Category	Product
206	182004189	Company D	Bangkok	www.business.co.th	0851245980	E-Commerce	Website
207	202001064	Company E	Phuket	www.pkdev.com	0946964468	Internal Mgmt	Smart Farm
224	202001278	Company F	Bangkok	www.prof.co.th	0661344114	Production Mgmt	AR VR

Figure 1. Comparison of digital provider data storage between Organization A and Organization B.

Without structured semantic relationships, organizations suffer from a lack of semantic interoperability, which extends beyond simple formatting differences in databases [12]. While Figure 1 demonstrates structural variations in data storage, real-world Dx project environments generate a significantly more complex array of unstructured information, including multimedia artifacts such as training videos, demonstration recordings, project images, and financial reports [3,4]. These resources are often siloed across different repositories and departments, making the manual extraction of insights—such as identifying historical integration challenges or specific technology performance—a process highly prone to human error [10].

Consequently, there is a requirement for a consistent framework to model this multi-format project knowledge through semantic ontology, which can unify diverse data types under a single machine-readable structure based on clearly defined concepts [13]. While ontology-based approaches have demonstrated potential in improving knowledge inte-

gration and semantic search [14], their practical adoption within the Dx domain remains limited [7]. As noted by [15], current semantic applications are often characterized by low corporate adoption rates and tend to be confined to isolated technical silos, highlighting the need for a more comprehensive core ontology that can manage the operational reality of heterogeneous project artifacts. Critically, no existing ontology provides both the strategic-level vocabulary (e.g., transformation phases, strategic dimensions) and the operational-level structure (e.g., multimedia artifact types, technology products) necessary to fully index and analyze Dx project portfolios.

To address this gap, this research proposes a semantic ontology framework to manage Dx project knowledge. The central component is Dx-Onto, a domain-specific core ontology implemented in OWL and designed to model essential project elements, including phases, strategic dimensions, and multimedia artifacts. Dx-Onto unifies data from diverse formats, facilitating semantic search, comparative analysis of successful cases, and identification of key technologies. In summary, this paper makes the following contributions:

1. It proposes Dx-Onto, a core ontology for representing Dx project knowledge with modular coverage of strategic dimensions, project phases, and operational artifacts.
2. It demonstrates five query-driven use cases showing how Dx-Onto supports multimedia artifact retrieval, semantic filtering, and multidimensional analysis.
3. It conducts scalability experiments using synthetic datasets to evaluate query performance under increasing data loads.
4. It presents a comparative benchmark evaluation against the Core Ontology of Organisational Transformation (COOT) [16] using a real-world heterogeneous corpus of Thai Dx documentation, demonstrating Dx-Onto's superior domain fit and unique analytical capabilities.

2. Background

This section introduces the foundational concepts that underpin the proposed framework, providing the necessary context for understanding the design rationale and scope of Dx-Onto.

2.1. Semantic Ontology and Ontology Types

In the context of knowledge engineering, an ontology is defined as "an explicit specification of a conceptualization" [17]. It provides a formal, machine-readable vocabulary of concepts, properties, and relationships within a domain, enabling both humans and software agents to share a common understanding of its structure [18]. The Web Ontology Language (OWL), a W3C standard built upon RDF and the Semantic Web vision [19,20], is the predominant formalism for representing ontologies, supporting class hierarchies, property restrictions, and logical inference.

Ontologies are generally categorized by their level of generality [18]: (1) Upper (or foundational) ontologies describe very general concepts (e.g., object, event, process) that are domain-independent, providing a common grounding for cross-domain interoperability. (2) Domain ontologies specialize upper-level concepts for a particular field, such as medicine, engineering, or digital transformation. (3) Application ontologies are further specialized for a specific task or use case within a domain. Dx-Onto is designed as a core domain ontology for digital transformation, aiming to provide a reusable, shared vocabulary that is specific enough to capture Dx project knowledge yet general enough to be applicable across diverse organizational contexts.

2.2. Digital Transformation: Stages and Dimensions

Digital transformation refers to the integration of digital technologies into business processes to enhance performance, efficiency, and customer value. This process fundamentally relies on strategic alignment, emphasizing the co-evolution of business strategy and information technology infrastructure [21]. It occurs in three evolutionary stages: digitization (conversion of analog data into digital form), digitalization (using digital tools to optimize workflows), and digital transformation (redefining business models using technologies like AI and blockchain) [2]. These evolutionary stages can be comprehensively examined through the Technology–Organization–Environment (TOE) framework [22,23] and the Dynamic Capabilities theory [24,25]. From this perspective, recent literature emphasizes that Dx is not merely the technical acquisition of digital tools, but a continuous strategic reconfiguration of an organization's resources to sense, seize, and transform in response to rapidly changing digital environments [25].

Successful Dx requires alignment with nine key dimensions: structural, informational, environmental, security, quality, financial, cultural, innovation, and participative aspects [5]. To navigate these complexities, scholars have proposed Digital Maturity Models (DMMs) to assess organizational readiness, arguing that leadership and digital capabilities must advance in tandem [26,27]. This multidimensional view is consistent with prior large-scale reviews, which emphasize that Dx success depends not only on technology adoption but also on organizational capabilities, governance structures, and knowledge integration across business units [7,9,28].

2.3. Knowledge Management in Digital Transformation

Knowledge management (KM) encompasses the processes of creating, sharing, using, and managing knowledge assets within an organization [29,30]. In the context of Dx, KM faces several acute challenges. As organizations scale their digital capabilities, the sheer volume and heterogeneity of project artifacts—ranging from strategic plans and financial spreadsheets to multimedia recordings—often outpace their knowledge management capacities [31]. Key challenges include: (1) knowledge fragmentation across departmental silos, where valuable insights remain inaccessible for cross-project reuse; (2) format heterogeneity, where structured databases, semi-structured spreadsheets, and unstructured documents must coexist; and (3) semantic ambiguity, where the same concept may be represented differently across teams, hindering consolidation [8,32]. These challenges motivate the use of ontology-based approaches to provide a formal, shared semantic backbone for Dx knowledge.

2.4. Challenges of Ontology Engineering

While ontology-based approaches offer significant advantages for knowledge integration, the ontology engineering process itself presents several recognized challenges that may affect practical implementation [33,34]. First, the knowledge acquisition bottleneck—the difficulty of eliciting tacit domain knowledge from experts and formalizing it into an ontology—remains a persistent challenge [35]. Second, ontology development requires specialized expertise in formal modeling languages (e.g., OWL, SPARQL) and conceptual design, which may not be readily available within project teams [15]. Third, maintenance and evolution of ontologies over time is a non-trivial concern, as domain knowledge evolves and new concepts emerge [36]. Finally, scalability of ontology reasoning and query performance must be carefully managed as the knowledge base grows [34].

To mitigate these challenges, this study adopts a modular, iterative design approach with clearly defined competency questions. The use of well-established OWL standards ensures tooling support and interoperability, while the modular architecture allows in-

cremental extension without requiring a complete redesign. The scalability concern is explicitly addressed through the experimental evaluation presented in Section 7.

3. Related Works

This section reviews existing ontology-based approaches for knowledge management and digital transformation, identifies current research gaps, and provides a systematic comparison to justify the necessity of Dx-Onto.

3.1. Ontology-Based Knowledge Management

Ontology-based KM has been extensively studied over the past two decades, especially in the context of the Semantic Web and enterprise systems [37]. It employs formal ontologies to structure and share knowledge in a machine-readable form, enabling efficient retrieval and reuse across organizations.

In practice, domain experts develop ontologies by defining key concepts, attributes, and relationships. Knowledge assets (e.g., documents, reports, data records) are then annotated accordingly, enabling semantic queries across linked data. For instance, a query on a specific component retrieves related documents through ontology relationships [38]. A key advantage is semantic search and inference—users can query concepts, and the system leverages hierarchies and graph links to retrieve directly or indirectly related information. For example, a search for "mobile payment projects" may yield relevant documents even without exact keywords, provided the ontology encodes the relationship [39]. Ontology-based KM also supports reasoning; axioms or business rules enable inference, such as identifying high-risk projects with unfamiliar technologies [40,41].

These benefits have been demonstrated across multiple domains, including healthcare, finance, and government [42]. In Industry 4.0, they support knowledge formalization across manufacturing and construction [43]. Recent surveys on knowledge graphs confirm the effectiveness of ontology-driven knowledge repositories in enterprise environments, particularly for integrating heterogeneous data sources [44].

Project management is another major area of application. Projects like ProjectCO and SPMO utilize ontologies derived from PMBOK to model project elements [45,46]. Other efforts target risk management [47] and strategic alignment [48]. However, traditional approaches often struggle with the heterogeneity of artifacts in Dx projects. Table 1 summarizes the differences between traditional data management and the ontology-based approach proposed in this study.

Table 1. Comparison between Traditional Approaches and Ontology-Based Approaches in Dx Context.

Feature	Traditional Approaches (DBMS/File Systems)	Ontology-Based Approach (Dx-Onto)
Data Structure	Rigid schemas (SQL) or unstructured hierarchies (Folders).	Flexible graph structure; easily extensible classes/properties.
Query Capability	Keyword-based; relies on exact string matching.	Semantic-based; understands context, relationships, and inference.
Interoperability	Low; siloed data with specific formats (Excel, PDF).	High; unified vocabulary linking heterogeneous artifacts.
Context Awareness	Limited; metadata is often detached from content.	Rich; links artifacts to project phases, KPIs, and stakeholders.

3.2. Existing Ontologies for Digital Transformation

Existing Dx ontologies often define high-level concepts such as "Digital Capability" or "Business Process" to model strategic outcomes. For instance, the Core Ontology of Organisational Transformation (COOT) by [16] establishes a foundational reference model for conceptualizing the structural evolution of business components during transformation. COOT defines 22 classes covering transformation drivers, stakeholders, resources, and goals, providing a general-purpose vocabulary for organizational change. Similarly, frameworks like DT&I-BPM-Onto illustrate how ontologies can align transformation initiatives with business processes [49].

However, the practical adoption of these semantic models often faces a bottleneck due to the complexity of the engineering process itself. To address this, Ref. [15] evaluated agile Ontology Engineering Methodologies (OEMs), such as UPONLite and SAMOD, highlighting that successful digitalization of knowledge management requires flexible, iterative workflows that match a firm's internal expertise.

While foundational and methodological studies provide the "how-to," recent specialized research has shifted toward technical and industrial domains. In the industrial sector, Ref. [50] demonstrated how knowledge graphs and upper ontologies bridge the gap between stakeholders and fragmented data sources, ensuring semantic interoperability across the asset lifecycle. This is further mirrored in the Architecture, Engineering, and Construction (AEC) sector, where Ref. [14] introduced an ontology-driven framework to support Material Information Modelling (MIM). Their work emphasizes the use of Linked Data and FAIR principles (Findable, Accessible, Interoperable, and Reusable) to overcome data exchange deficiencies in building performance simulation.

Despite these advancements, most existing approaches remain either purely descriptive at a high organizational level [16] or restricted to narrow technical silos like asset maintenance [51] or material science [14]. Prior work rarely addresses the operational reality of managing the heterogeneous artifacts—ranging from administrative spreadsheets and PDF reports to project-specific multimedia—generated across the various phases of a Dx journey.

3.3. Research Gap and Comparative Analysis

Recent literature indicates that the primary bottleneck in modern semantic applications is not conceptual modeling, but operational scalability and structural integration within heterogeneous enterprise environments [52,53]. While foundational ontologies address high-level business capabilities, they often lack the specific mechanisms to represent operational data—such as integrating raw spreadsheets, multimedia files, and performance metrics into a unified Semantic Web framework [54]. Furthermore, many current semantic approaches are confined to isolated academic pilots or specialized technical silos (e.g., exclusively for BIM or manufacturing), lacking a generalized schema to handle the diverse administrative artifacts generated throughout a Dx lifecycle [55].

To clearly delineate the contributions and limitations of prior work, Table 2 presents a systematic comparison of key ontological features across existing frameworks and the proposed Dx-Onto. This comparison highlights the specific conceptual and structural gaps that Dx-Onto addresses.

As shown in Table 2, while COOT provides a broad organizational transformation vocabulary and DT&I-BPM-Onto contributes business process alignment, neither framework captures the Dx-specific vocabulary of transformation phases, strategic dimensions, technology ecosystems, or heterogeneous multimedia artifacts that are essential for operational project knowledge management. To date, no widely accepted core ontology systematically supports comprehensive knowledge reuse and cross-project analysis by

unifying strategic dimensions with these diverse digital resources. This gap motivates the proposed Dx-Onto, which serves as a pragmatic semantic layer designed for scalable and reliable Dx knowledge management.

Table 2. Comparative analysis of ontological features across existing frameworks and the proposed Dx-Onto. Checkmark (✓) indicates the feature is supported; dash (–) indicates not supported.

Ontological Feature/Concept	COOT [16]	DT&I-BPM-Onto [49]	Dx-Onto (Proposed)
Transformation Drivers/Goals	✓	✓	✓
Stakeholders/Actors	✓	✓	✓
Organizational Resources	✓	–	✓
Business Processes	–	✓	–
Dx Phases (Digitization/Digitalization/DT)	–	–	✓
Strategic Dimensions (9 subclasses)	–	–	✓
Technology Products & Providers	–	–	✓
KPIs/Performance Metrics	–	–	✓
Multimedia Artifact Types (PDF, Image, Video)	–	–	✓
Project Lineage (Pre/Post-Project)	–	–	✓
OWL Implementation	✓	Partial	✓
Empirical Benchmark with Real Data	–	–	✓

4. Research Methodology

This study adopts an ontology engineering methodology to design, implement, and evaluate Dx-Onto. Before describing the process, we briefly justify the choice of ontology engineering over alternative knowledge integration approaches.

Several alternative approaches exist for integrating heterogeneous project data, including relational database schemas, document-oriented NoSQL stores, and unstructured knowledge graphs without formal ontological grounding. While relational databases excel at structured query performance, they impose rigid schemas that cannot accommodate the format heterogeneity of Dx artifacts (PDFs, images, spreadsheets) without extensive preprocessing [52]. NoSQL stores offer schema flexibility but lack the formal semantic relationships and inference capabilities required for cross-project knowledge discovery. Pure knowledge graphs without an ontological backbone can represent arbitrary relationships but provide no domain-specific vocabulary or reasoning support [19]. Ontology engineering, by contrast, offers: (1) a formal, reusable vocabulary with defined class hierarchies and property restrictions; (2) support for logical inference and semantic querying via SPARQL; and (3) interoperability through W3C standards (OWL, RDF). These advantages make it the most suitable approach for structuring the multidimensional, heterogeneous knowledge generated across Dx project lifecycles [18,36].

We acknowledge that ontology engineering requires specialized knowledge in formal modeling and Semantic Web technologies, which may not be available in all organizational contexts [15]. To mitigate this, Dx-Onto was designed with a modular architecture that separates concerns (strategic dimensions, project phases, operational artifacts) into manageable sub-modules. This modularity enables incremental development and lowers the barrier for teams with varying levels of ontology engineering experience. Additionally, the use of established standards (OWL, Protégé) and comprehensive documentation facilitates knowledge transfer and long-term maintenance [33].

The development process follows a five-phase lifecycle: (1) Requirement Analysis, (2) Ontology Design & Implementation, (3) Population, (4) Scalability Evaluation, and (5) Comparative Benchmark Evaluation, as depicted in Figure 2.

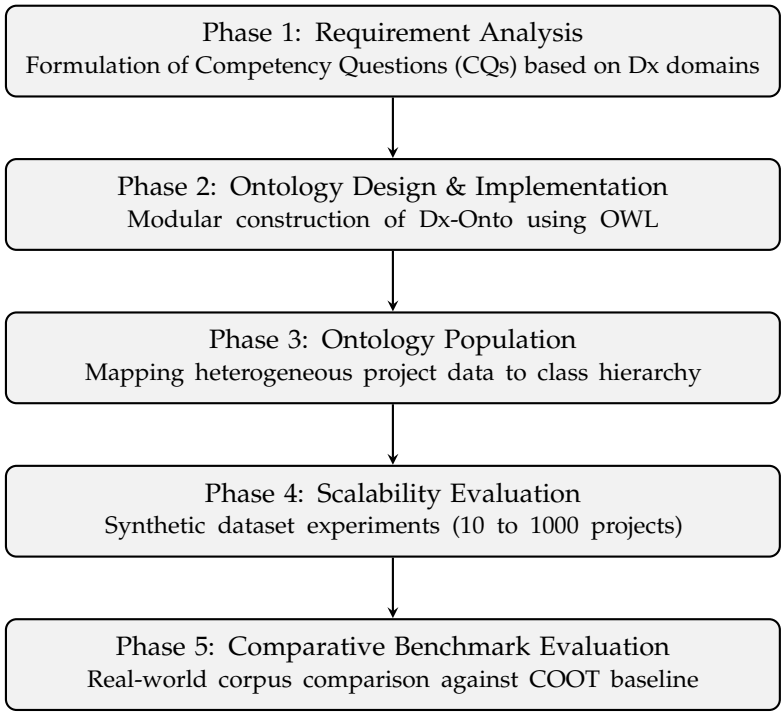


Figure 2. The five-phase ontology engineering methodology adopted for the development and evaluation of Dx-Onto.

4.1. Phase 1: Requirement Analysis and Competency Questions

Requirements were elicited from domain literature, standard Dx principles, and empirical observations from public and private sector projects. A key input was the motivating scenario as described in Section 1, which highlighted the challenges of data fragmentation. Based on these inputs, a set of Competency Questions (CQs) was formulated to define the scope and querying capabilities of the ontology (e.g., "Which projects used Technology X and achieved KPI Y?").

4.2. Phase 2: Ontology Design and Implementation

The ontology design follows a modular approach. The core domain ontology (Dx-Onto) was constructed to represent specific Dx concepts (Phases, Dimensions, KPIs, Artifacts). The ontology was implemented using OWL (Web Ontology Language) to ensure semantic interoperability and support for reasoning.

4.3. Phase 3: Ontology Population

To validate the ontology, instances were populated based on the motivating scenario and real-world project structures. This phase involved mapping heterogeneous data sources—such as spreadsheet columns and report metadata—into the ontology's class hierarchy (e.g., mapping a "budget" column to `dx:hasBudget`).

4.4. Phase 4: Scalability Evaluation

To assess the scalability and efficiency of Dx-Onto, an experimental evaluation was conducted using synthetic datasets. These datasets were algorithmically generated to mimic the structural complexity and statistical distribution of real-world projects, ranging from small-scale (10 projects) to large-scale (1000 projects) scenarios. This approach allows

for a controlled evaluation of query performance (Median and p95 execution times) under increasing data loads.

4.5. Phase 5: Comparative Benchmark Evaluation

To address the need for empirical validation with real-world data and comparison against an existing baseline, a comparative benchmark was conducted. Dx-Onto was evaluated against the Core Ontology of Organisational Transformation (COOT) [16] using a heterogeneous corpus of real Thai digital transformation documentation. An identical LLM-based extraction pipeline was applied to both ontologies to ensure a fair comparison. The detailed setup, results, and discussion of this benchmark are presented in Section 8.

5. Digital Transformation Ontology (Dx-Onto) Design

This section presents the design of the Digital Transformation Ontology (Dx-Onto) (https://github.com/SareeyaBen/DxOntology, accessed on 17 March 2026), which serves as the core semantic framework for managing knowledge in Dx projects. The primary goal of Dx-Onto is to provide a unified schema that interconnects heterogeneous project artifacts, such as documents, spreadsheets, and multimedia, with the strategic dimensions of digital transformation. To ensure comprehensive coverage of the digital ecosystem, Dx-Onto is designed with modularity in mind, focusing on the project lifecycle while maintaining the capability to represent essential contextual elements such as technology vendors and products within the knowledge graph.

5.1. Core Structure of Dx-Onto

The architecture of the Dx-Onto schema is illustrated in Figure 3.

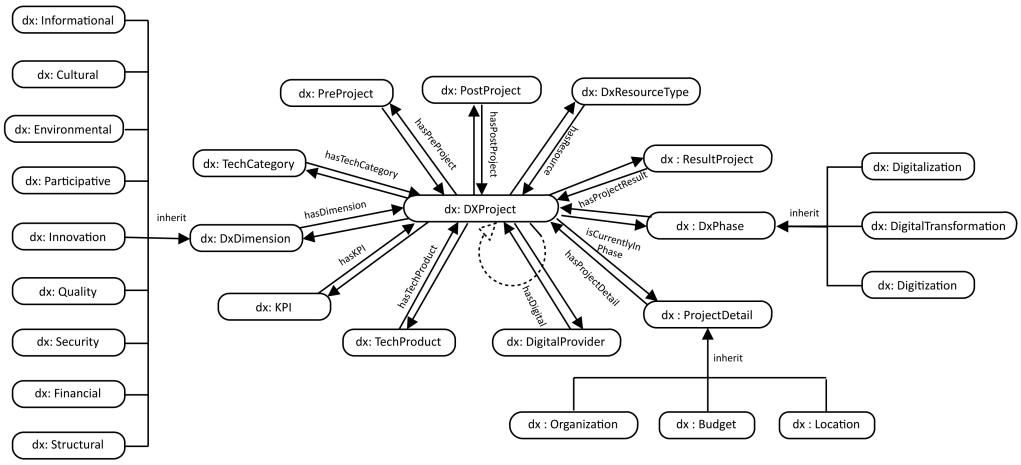


Figure 3. Dx-Onto: The Core Ontology for Digital Transformation Knowledge Management. In Figure 3, primary class relationships are distinguished by solid lines, while cyclical relationships (e.g., connections between sequential projects) are indicated by dashed lines.

The ontology is structured around the central class `dx:DxProject`, which acts as the semantic hub connecting all critical aspects of a digital transformation initiative. This core entity is linked to eleven primary classes through specific object properties, ensuring a comprehensive representation of project knowledge. These components are categorized as follows:

- Strategic Dimensions (`dx:DxDimension`): To capture the multi-faceted nature of transformation, this class is linked via the `hasDimension` property. It is further specialized into nine distinct subclasses representing key organizational aspects: `dx:Structural`, `dx:Financial`, `dx:Security`, `dx:Quality`, `dx:Innovation`,

`dx:Participative`, `dx:Environmental`, `dx:Cultural`, and `dx:Informational`. This granularity allowing organizations to categorize and analyze projects based on their primary strategic impact.

- Project Lifecycle (`dx:DxPhase`): The temporal progress of a project is tracked using the `isCurrentlyInPhase` property. The `dx:Phase` class distinguishes between the three fundamental stages of digital maturity: `dx:Digitization` (converting analog to digital), `dx:Digitalization` (process optimization), and `dx:DigitalTransformation` (business model reinvention).
- Operational Details (`dx:ProjectDetail`): This class encapsulates the practical constraints and context of a project via the `hasProjectDetail` property. It includes specific subclasses for `dx:Budget` (financial resources), `dx:Location` (geographical or departmental implementation site), and `dx:Organization` (the responsible entity).
- Knowledge Artifacts (`dx:DxResourceType`): A critical component for solving data fragmentation, this class is linked via `hasResource`. It represents the heterogeneous files generated during a project (e.g., PDFs, images, spreadsheets), enabling the system to map unstructured files directly to structured project concepts.
- Technology Ecosystem: To model the technological landscape, the ontology connects projects to specific tools and vendors through three classes: `dx:DigitalProvider` (via `hasDigitalProvider`), `dx:TechProduct` (via `hasTechProduct`), and `dx:TechCategory` (via `hasTechCategory`). These links allow users to trace which vendors and technologies are driving specific initiatives.
- Performance and Results: To evaluate success, the ontology defines `dx:KPI` (via `hasKPI`) for quantitative metrics and `dx:ResultProject` (via `hasProjectResult`) for qualitative outcomes (e.g., "Success" or "Failure").
- Project Lineage: Finally, to support project continuity and dependency tracking, the ontology models the relationships between sequential initiatives using `dx:PreProject` (via `hasPreProject`) and `dx:PostProject` (via `hasPostProject`).

5.2. Application Scenarios

To demonstrate the practical utility of Dx-Onto, we present three specific modeling scenarios that address the knowledge management challenges identified in the introduction.

5.2.1. Scenario 1: Multimedia Data Retrieval

One of the primary challenges in Dx management is retrieving unstructured multimedia data scattered across different folders. Figure 4 illustrates how Dx-Onto models this scenario.

In this model, projects such as SatunTravel and SongklaCoding are linked to their respective files via the `hasResource` property. For example, Satuntravel.jpeg is semantically defined as an Image type. This structure allows users to execute precise queries, such as "Retrieve all image files related to tourism projects," overcoming the limitations of traditional keyword-based file search.

Multimedia Ingestion Process. We clarify that the mapping of unstructured multimedia artifacts into the ontology follows a semi-automated pipeline. File-level metadata (e.g., file name, extension, MIME type, creation date) is extracted automatically during ingestion, and the file type classification (e.g., `ImageType`, `DocumentType`, `VideoType`) is determined programmatically based on file extension and MIME type mapping. For richer semantic annotation—such as associating a file with a specific project phase or strategic dimension—the current implementation relies on structured metadata provided by project administrators during the upload process. Future iterations of the framework envision

deeper automated extraction using computer vision for image analysis and NLP-based text extraction from PDFs to further reduce manual effort.

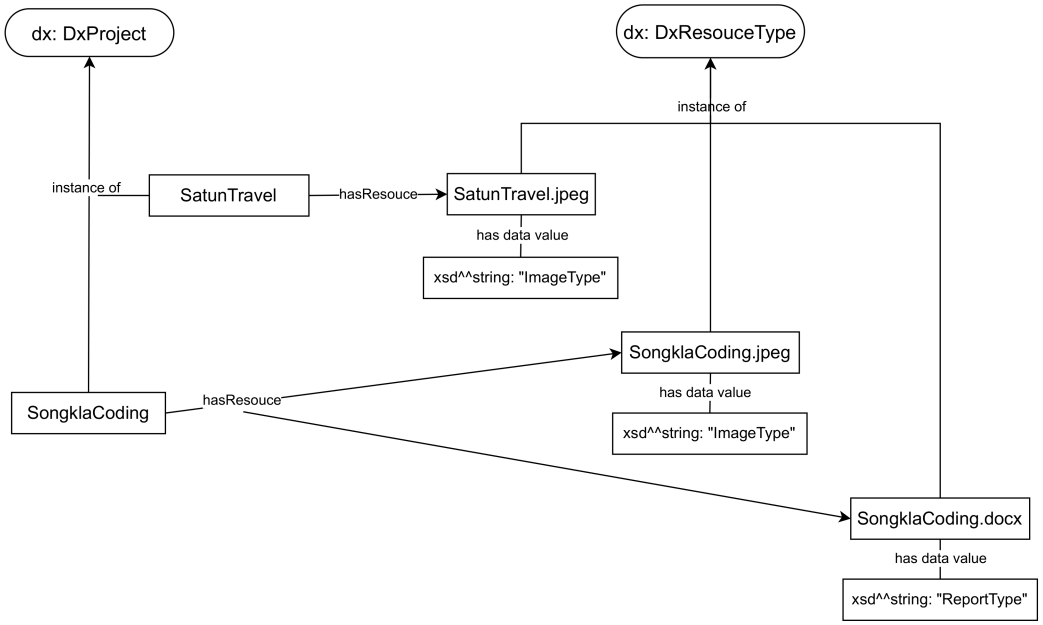


Figure 4. Using Dx-Onto to model and retrieve multimedia data.

5.2.2. Scenario 2: Semantic Information Discovery

Beyond file retrieval, decision-makers often need to find projects based on complex criteria, such as technology type and success status. Figure 5 demonstrates this capability.

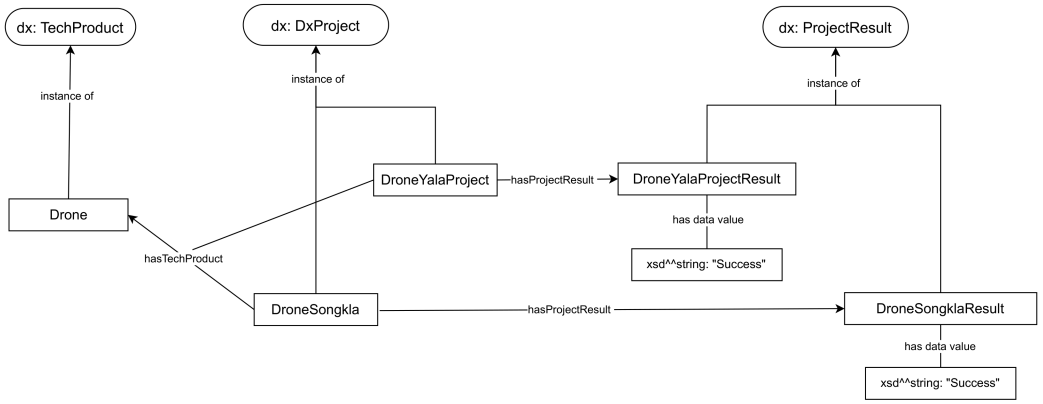


Figure 5. Using Dx-Onto for semantic discovery of successful projects.

Here, the ontology links DroneYalaProject and DroneSongkla to the Drone technology product and a Success result. This semantic linkage enables the system to answer competency questions like "Identify all successful drone projects in Southern Thailand," which would otherwise require manual cross-referencing of multiple reports.

5.2.3. Scenario 3: Multidimensional Process Analysis

The final scenario focuses on strategic analysis, specifically identifying which transformation dimensions contribute to project profitability. Figure 6 depicts this modeling approach.

In this example, tourism projects like Travel_Bus are linked to specific dimensions (Finance, Innovation) and KPIs (Profit increase > 10%). By structuring data in this way, Dx-Onto enables researchers to correlate specific transformation activities with financial

outcomes. This capability is essential for organizations aiming to replicate the success of high-performing projects in future initiatives.

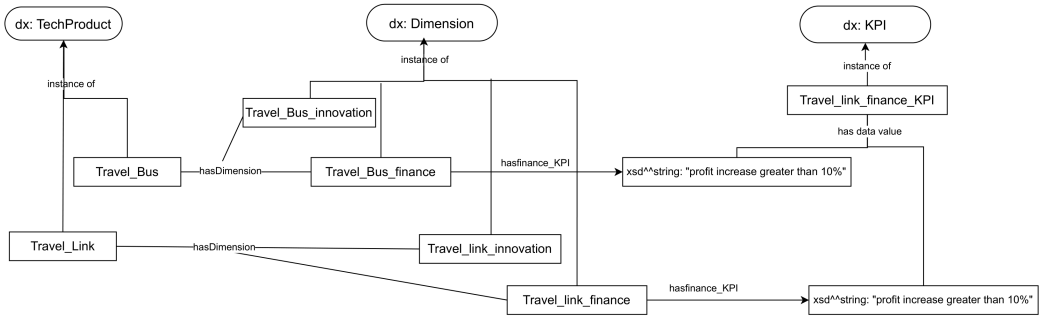


Figure 6. Using Dx-Onto to analyze dimensions of profitable tourism projects.

While the conceptual models presented above illustrate the semantic expressiveness of Dx-Onto, the practical utility of the framework relies on its ability to support machine-readable queries. To demonstrate how these semantic structures are operationalized, the next section translates the motivating scenarios into concrete SPARQL queries. By mapping the competency questions defined in the methodology to executable code, we validate that Dx-Onto can effectively support the retrieval of heterogeneous artifacts, cross-project discovery, and multidimensional analysis as intended.

6. Example Query Use Cases on Dx-Onto

With the ontology constructed and populated with representative data, we validated its capability by executing SPARQL queries against the knowledge base. This systematic evaluation phase is specifically designed to ensure semantic interoperability and verify the precise retrieval of heterogeneous project artifacts.

The following use cases demonstrate how Dx-Onto answers the competency questions defined in the methodology, ranging from simple artifact retrieval to complex multi-dimensional analysis. By mapping these natural language requirements into formal, machine-readable syntax, these scenarios practically illustrate the framework's capacity to bridge the gap between high-level strategic dimensions and low-level operational resources.

6.1. Use Case 1: Multimedia Artifact Retrieval

Requirement: Retrieve image files from the "Satun Travel" project and both Word/Image files from the "Coding School" project.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?projectName
        ?fileType
        ?fileName

WHERE {
  ?project a dx:DxProject ;
           dx:projectName ?projectName ;
           dx:hasResource ?resource~.

  ?resource a dx:DxResourceType ;
            dx:hasFileType ?fileType ;
            dx:hasFileName ?fileName~.
```

```
  FILTER (
    (?projectName = "Satun travel project" &&
     ?fileType = "image")
    ||
    (?projectName = "Coding School project" &&
     (?fileType = "image" ||
      ?fileType = "word"))
)
}
```

This query traverses the graph from `dx:DxProject` to `dx:DxResourceType` via the `dx:hasResource` property. It retrieves the file name and type, applying a nested filter to select specific artifacts based on project names. Specifically, the SPARQL pattern utilizes logical && (AND) and || (OR) operators within the FILTER clause to precisely isolate image files associated with the "Satun travel project", while simultaneously retrieving both image and word processing documents linked to the "Coding School project".

By executing this single declarative query instead of manually navigating through disjointed directory systems, the framework significantly reduces retrieval latency. This demonstrates Dx-Onto's ability to consolidate heterogeneous file formats under a unified semantic structure.

6.2. Use Case 2: Geospatial and Technology-Based Discovery

Requirement: Identify successful drone projects in Yala and Songkla provinces.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?projectName ?province ?techProductName ?kpiName

WHERE {
  ?project a dx:DxProject ;
           dx:projectName ?projectName ;
           dx:hasProjectDetail ?detail ;
           dx:hasTechProduct ?tech ;
           dx:hasProjectResult ?result ;
           dx:hasKPI ?kpi~.

  ?detail dx:location ?province .
  ?tech dx:productName ?techProductName .
  ?kpi dx:kpiName ?kpiName .
  ?result dx:isProjectResult "successful"~.

  FILTER(CONTAINS(LCASE(?province), "yala") ||
         CONTAINS(LCASE(?province), "songkla")) .

  FILTER(CONTAINS(LCASE(?techProductName), "drone")) .
}
```

This query links multiple concepts: Location (via `dx:ProjectDetail`), Technology (via `dx:TechProduct`), and Outcome (via `dx:ResultProject`). It filters for projects in specific provinces that utilize "drone" technology and have achieved a "successful" status. This illustrates how Dx-Onto supports complex semantic filtering to discover niche projects.

6.3. Use Case 3: Strategic Dimension Analysis

Requirement: Identify which transformation dimensions are critical to profitable tourism projects.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT DISTINCT ?projectName ?dimensionName
WHERE {
  ?project a dx:DxProject ;
           dx:projectName ?projectName ;
           dx:hasDimension ?dimension ;
           dx:hasProjectResult ?result ;
           dx:hasProjectDetail ?detail~.

  ?dimension dx:dimensionDescription ?dimensionName .
  ?result dx:isProjectResult "successful" .
  ?detail dx:budget ?budget .
  FILTER(CONTAINS(LCASE(?projectName), "tourism")) .
  FILTER(CONTAINS(LCASE(?budget), "profit")) .
}
```

This query correlates financial success with strategic dimensions. It filters for successful tourism projects that are profitable and retrieves their associated dimensions (e.g., "Innovation", "Cultural"). This allows stakeholders to understand which strategic pillars drive profitability in specific sectors.

6.4. Use Case 4: Advanced Maturity and Innovation Analysis

Requirement: Retrieve high-maturity projects that are currently in the "Digital Transformation" phase (the highest maturity level), classified as successful, and driven by the "Innovation" dimension.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?projectName ?kpiValue
WHERE {
  ?project a dx:DxProject ;
           dx:projectName ?projectName ;
           dx:isCurrentlyInPhase ?phase ;
           dx:hasProjectResult ?result ;
           dx:hasDimension ?dimension ;
           dx:hasKPI ?kpi .
  FILTER regex(str(?phase), "DigitalTransformation", "i") .
  ?result dx:isProjectResult "successful" .
  ?dimension dx:dimensionDescription "Innovation" .
  ?kpi dx:hasValue ?kpiValue .
}
```

This advanced query targets projects at the peak of the transformation lifecycle. By filtering for the specific `dx:DigitalTransformation` phase (distinct from earlier Digitization/Digitalization phases) and intersecting it with the `dx:Innovation` dimension, this query isolates high-impact, mature initiatives. It serves as a powerful tool for policymakers to identify "role model" projects for scaling.

6.5. Use Case 5: Comparative Analysis (Lesson Learned)

Requirement: Identify pairs of projects that utilized the same technology product but achieved different results (one successful, one failed) to facilitate comparative analysis.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?techName ?successProject ?failProject
WHERE {
  ?tech a dx:TechProduct ;
        dx:productName ?techName .
  ?projA a dx:DxProject ;
         dx:projectName ?successProject ;
         dx:hasTechProduct ?tech ;
         dx:hasProjectResult ?resA .
  ?resA dx:isProjectResult "successful" .
  ?projB a dx:DxProject ;
         dx:projectName ?failProject ;
         dx:hasTechProduct ?tech ;
         dx:hasProjectResult ?resB .
  ?resB dx:isProjectResult "failure" .
}
```

This use case highlights the Knowledge Management capability of Dx-Onto. By identifying projects that employed identical technologies but yielded contrasting outcomes, organizations can perform comparative analysis to understand root causes (e.g., whether the failure was due to the technology itself or the implementation strategy). This directly supports the "Lesson Learned" objective of the framework.

7. Experimental Performance Evaluation

While the previous sections demonstrated the semantic expressiveness of Dx-Onto, ontology-based systems are often constrained by computational overhead compared to traditional databases. To assess the practical viability of Dx-Onto, we conducted controlled scalability experiments to measure query response times under increasing data loads.

7.1. Experimental Setup

To ensure a rigorous stress test, we generated synthetic datasets ranging from 10 to 1000 projects. Synthetic data allowed for systematic variation of volume while preserving the complex semantic structure of real-world Dx projects. For the technical implementation, the ontology was loaded into an in-memory graph using RDFLib (https://rdflib.readthedocs.io/en/stable/index.html, accessed on 10 February 2026), a widely adopted and industry-standard Python 3.12 library in the Semantic Web community. This prominent open-source library was selected due to its comprehensive support for W3C standards (https://www.w3.org/RDF/, accessed on 10 February 2026), including RDF 1.1 and SPARQL 1.1, which enables the efficient graph manipulation and serialization required to execute the complex semantic queries in our evaluation setup.

We evaluated two distinct query patterns representing typical user workloads:

1. Broad Document Retrieval: A data-intensive query that retrieves all multimedia artifacts associated with a specific project.
2. Cross-Project Semantic Filtering: A complex analytical query that filters projects based on multi-dimensional criteria (Location + Technology + Result).

For each dataset size, queries were executed over 20 independent trials, recording SD, Median and 95th Percentile (p95) execution times.

7.2. Experiment 1: Broad Document Retrieval

In this first experiment, we focus on a data-intensive scenario: retrieving all multimedia artifacts (e.g., images, documents) associated with a single project. This simulates a common user action where a stakeholder browses a project's digital repository. Since a project can have multiple linked resources, this query tests the system's ability to traverse one-to-many relationships efficiently.

The SPARQL query used for this test is shown below. It matches a specific project (e.g., "Project_1") and retrieves every resource linked via the `dx:hasResource` property.

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?projectName ?fileType ?fileName WHERE {
  ?p a dx:DxProject ;
     dx:projectName ?projectName ;
     dx:hasResource ?res .
  ?res a dx:DxResourceType ;
       dx:hasFileType ?fileType ;
       dx:hasFileName ?fileName .
  FILTER(STRSTARTS(?projectName, "Project_1"))
  FILTER(?fileType = "image")
}
```

The performance results obtained from executing this query across different dataset sizes are summarized in Table 3.

Table 3. Performance Results for Broad Document Retrieval.

Projects	Median (ms)	SD (ms)	p95 (ms)
10	7.64	0.30	8.14
50	51.49	0.74	52.70
100	174.79	3.11	179.90
200	656.92	20.67	690.93
500	4009.54	109.16	4189.10
1000	15,545.25	436.01	16,262.48

As observed in Table 3, the execution time exhibits a non-linear growth relative to the dataset size. For small to medium-scale repositories (up to 100 projects), the system maintains high efficiency with a median latency below 200 ms. The low Standard Deviation (SD) across these scales (ranging from 0.30 to 3.11 ms) indicates high system stability and consistency in query processing. However, as the project volume increases to 1000 instances, the median retrieval time reaches 15,545.25 ms. The p95 latency of 16,262.48 ms further reveals that even in the worst-case scenarios, the performance remains within a narrow margin of the median, suggesting that the computational overhead is predictable rather than erratic. This trend is consistent with the structural complexity of RDF triple-store reasoning during broad-spectrum artifact retrieval. While the increased latency at 1000 projects represents a trade-off for ensuring semantic completeness, the predictable nature of the SD and p95 values confirms that Dx-Onto provides a reliable foundation for large-scale data retrieval, which can be further optimized through triple-store indexing in production environments.

7.3. Experiment 2: Cross-Project Semantic Filtering

Moving beyond simple retrieval, the second experiment evaluates the ontology's analytical capabilities. Here, we perform a cross-project search to identify initiatives that meet a complex set of criteria: located in specific provinces, utilizing drone technology, and achieving a successful outcome. This query is computationally more demanding in terms of logic because it requires joining multiple nodes (Project, Detail, Technology, Result) and applying string filters.

The following SPARQL query demonstrates this semantic filtering process:

```
PREFIX dx: <https://dxonto.105app.site/dx#>
SELECT ?projectName ?province WHERE {
  ?p a dx:DxProject ;
     dx:projectName ?projectName ;
     dx:hasProjectDetail ?detail ;
     dx:hasTechProduct ?tech ;
     dx:hasProjectResult ?result .
  ?detail dx:location ?province .
  ?tech dx:productName ?techProductName .
  ?result dx:isProjectResult "successful"~.

  FILTER(CONTAINS(LCASE(?techProductName), "drone"))
  FILTER(?province = "Yala" || ?province = "Songkhla" ||
         ?province = "Pattani" || ?province = "Satun")
}
```

Table 4 demonstrates superior scalability for semantic filtering tasks. Even at the 1000-project scale, the median execution time remains highly efficient at 2983.42 ms. Notably, the SD (48.91 ms) and p95 (3063.87 ms) show a remarkably tight distribution, proving that the semantic filtering logic in Dx-Onto is highly optimized for complex, multi-dimensional queries. This minimal variance between the median and p95 indicates that the ontology-driven search space reduction is effective, preventing the exponential performance degradation often seen in unoptimized graph queries. These results validate that Dx-Onto is not only scientifically sound but also practically viable for real-time decision support in large-scale enterprise ecosystems.

Table 4. Performance Results for Cross-Project Semantic Filtering.

Projects	Median (ms)	SD (ms)	p95 (ms)
10	8.15	0.30	8.64
50	18.04	0.27	18.48
100	44.61	5.00	52.83
200	141.39	4.94	149.52
500	771.75	28.60	818.79
1000	2983.42	48.91	3063.87

7.4. Scalability Evaluation Summary

The scalability experiments confirm that Dx-Onto is viable for core knowledge management tasks, particularly for complex, multi-dimensional queries (Experiment 2), which are the primary value proposition of this framework. While broad document retrieval (Experiment 1) shows latency at scale, this is a known trade-off in graph databases that can

be mitigated in real-world deployments through hybrid architectures (e.g., using search indices for artifacts and ontology for reasoning). Thus, Dx-Onto successfully balances semantic expressiveness with acceptable performance for decision-support queries.

8. Comparative Benchmark Evaluation

While the scalability experiments in Section 7 validated Dx-Onto's query performance under controlled synthetic conditions, a critical question remains: does the ontology effectively capture and structure knowledge from real-world Dx project data, and does it offer measurable advantages over existing alternatives? This section presents a head-to-head comparative benchmark between Dx-Onto and COOT [16], a published general-purpose ontology for organisational transformation, using a real-world heterogeneous corpus.

8.1. Experiment Setup

Objective: Evaluate whether Dx-Onto captures and structures knowledge more usefully than COOT when both are applied to the same real-world corpus of digital transformation documentation.

Corpus: Thailand Smart City Digital-Transformation documentation (https://data.go.th/dataset/, accessed on 2 June 2026), comprising CSV/Excel tables, PDF master plans, and one video (metadata only). The corpus is heterogeneous: structured project records, narrative master plans, company directories, product registries, and aggregate statistics.

Extraction Pipeline: An identical LLM-based pipeline was applied to both ontologies. Each file was text-extracted (PyMuPDF for PDF, openpyxl/xlrd for spreadsheets), chunked, and sent to an on-premise OpenAI-compatible endpoint running Qwen3.6-Plus (temperature 0). The model returned structured JSON, which was converted to RDF triples (RDFLib) under each ontology's schema. Values were stored bilingually (Thai/English). The only difference between the two runs was the target ontology and its prompt/triple-builder; the model, corpus, chunking, and parameters were identical.

Methodological Controls: To ensure a fair comparison, the following controls were applied:

- Duplicate-file removal before indexing (MD5), so repeated PDFs did not inflate counts.
- Global project deduplication by name, so a project appearing in multiple files contributes one entity set.
- Exclusion of schema examples: three example individuals shipped inside the Dx-Onto OWL file were excluded (COOT ships none).
- COOT harness parity: field-mapping bugs were fixed and missing prompt fields added so COOT could populate all classes its design supports (raising its reachable classes from 10 to 21 of 22).
- Dual class-usage metric: a class is counted as "used" if it is instantiated or referenced as a linked class IRI, applied symmetrically to both ontologies.

A separate counting pass using the same model (Qwen3.6-Plus) with a counting-only prompt enumerated distinct project mentions and rejected non-project content, providing an external reference for the extracted entity counts.

8.2. Results

Table 5 summarizes the comparative indexing performance of the two ontologies on the same corpus.

Because Dx-Onto defines dedicated vocabulary for transformation dimensions and maturity phases—concepts absent from COOT—it enables portfolio-level analytics that are structurally impossible under the baseline ontology. The dimension distribution across 134 extracted projects (not mutually exclusive) is: Innovation 41.0%, Participative 37.3%,

Informational 35.8%, Quality 34.3%, Environmental 25.4%, Structural 21.6%, Cultural 14.2%, Security 13.4%, Financial 10.4%. The maturity phase distribution is: Digitalization 56.7%, Digital Transformation 12.7%, Digitization 7.5%.

Table 5. Comparative indexing performance: Dx-Onto vs. COOT on a real-world Thai Dx corpus.

Metric	Dx-Onto	COOT
Indexed files (with $\geq$1 entity)	11	8
Total triples	5940	5078
Core entities (projects/transformations)	134	83
Total entities (all class instances)	1393	1384
Triple density (per core entity)	44.3	61.2
Thai labels	348	1361
English labels	345	1394
Ontology classes defined	27	22
Populated classes (instances only)	11	17
Used classes (instance or linked IRI)	23	17
Domain fit (used/defined)	85.2%	77.3%

The independent reference enumeration identified 489 distinct project mentions across the corpus, of which 361 (74%) came from two large community-contract registries. All files manually classified as non-project content (statistics, product registry, geographic mapping) returned 0 under the independent pass, in agreement with both ontology pipelines. The independent phase distribution—Digitalization 80.4%, Digital Transformation 17.4%, Digitization 1.0%—reproduces the same dominant-phase ordering as the Dx-Onto extraction, providing convergent evidence that the phase analytics reflect a genuine property of the portfolio rather than an extraction artifact.

8.3. Discussion of Comparative Results

Both ontologies extracted essentially the same total number of entities from identical input (1393 vs. 1384), so the comparison is not about extraction capacity. Dx-Onto found 61% more core projects (134 vs. 83) and drew from three more files (11 vs. 8), reflecting its broader fit to Dx content (industry data, technology catalogues) that COOT's narrower notion of "organisational transformation" did not admit. Conversely, COOT's higher triple density and approximately 4$\times$ more labels stem from its modeling style: it materializes every sub-concept (each stakeholder, resource, goal) as a separately labeled individual, whereas Dx-Onto stores attributes as compact literal properties and links controlled vocabularies as class IRIs. Neither style is intrinsically superior—the verbose form favors multilingual entity search; the compact form favors structured aggregate querying.

Under an instance-only count, Dx-Onto appears to populate only 11 of 27 classes, because it represents its nine dimensions and three phases as linked class IRIs rather than as instances—a valid OWL pattern. Counting a class as "used" when instantiated or referenced (applied symmetrically; COOT, which uses instances throughout, is unaffected) raises Dx-Onto to 23 of 27 (85.2%) versus COOT's 77.3%. The four remaining Dx-Onto classes are abstract grouping classes not intended for direct population.

The decisive advantage is structural, not quantitative. Dx-Onto's phase and dimension vocabularies enable portfolio-level analysis—such as identifying that the indexed portfolio is concentrated in the Digitalization phase (56.7%) and emphasizes Innovation, Participative, and Informational dimensions. These analyses are physically impossible

under COOT, which contains no phase or dimension vocabulary, regardless of extraction method or model capability. This demonstrates that the value of a domain-specific ontology lies not in how much knowledge it captures but in which classes of analytical question it renders answerable.

Overall, COOT achieves broad structural coverage of generic organisational concepts, while Dx-Onto provides comparable extraction volume plus transformation-specific analytics. For the task of indexing and analyzing a portfolio of digital-transformation projects, Dx-Onto's domain specificity is the deciding advantage.

9. Discussion

This section synthesizes the findings from both the scalability evaluation (Section 7) and the comparative benchmark (Section 8), and discusses broader implications, ontology engineering challenges, and limitations.

9.1. Scalability and Performance

The scalability experiments revealed a consistent distribution between Median, Standard Deviation (SD), and p95 values across both query types, suggesting that Dx-Onto's reasoning process is inherently predictable. The observed variance levels, even as dataset volume increases, indicate that the formal structural constraints effectively manage the computational complexities typically associated with graph queries. This stability provides a functional foundation for scalable knowledge management. These empirical results align with theoretical perspectives in recent digital transformation studies [7,16], which emphasize the necessity of structured semantic alignment to mitigate knowledge fragmentation in complex organizational environments.

9.2. Comparative Benchmark Insights

The comparative evaluation against COOT provided empirical evidence that the value of a domain-specific ontology lies not merely in extraction volume—where both ontologies performed comparably—but in the types of analytical questions it enables. Dx-Onto's dedicated vocabulary for transformation phases and strategic dimensions unlocks portfolio-level diagnostics (e.g., phase concentration, dimension emphasis) that are structurally unavailable in general-purpose transformation ontologies. This finding underscores the importance of domain-specific semantic modeling for operational knowledge management, beyond the high-level conceptual coverage that existing frameworks provide.

9.3. Ontology Engineering Challenges and Practical Implications

While the results demonstrate the analytical advantages of Dx-Onto, we acknowledge that ontology engineering remains a complex process with several unresolved challenges that may affect practical adoption [15,33]. The knowledge acquisition bottleneck required iterative consultation with Dx domain experts to refine class definitions and property constraints. Team expertise proved critical: the development team's familiarity with OWL semantics and SPARQL directly influenced the ontology's expressiveness and correctness. Organizations with limited Semantic Web experience may face steeper adoption curves, potentially affecting both the quality of the initial ontology and its long-term maintenance.

To address these concerns, Dx-Onto's modular architecture was specifically designed to separate independent sub-domains (phases, dimensions, artifacts, technology ecosystem) into loosely coupled modules. This design allows teams to incrementally extend or refine individual modules without requiring expertise in the entire ontology. Furthermore, the availability of the OWL file on GitHub, combined with standard Protégé tooling, reduces the barrier to entry for organizations seeking to adopt or customize the ontology for their specific contexts.

9.4. The Role of Ontologies in the LLM Era

The results obtained from Dx-Onto also address a critical scientific question in the contemporary AI landscape: whether Semantic Web technologies remain relevant in the era of Large Language Models (LLMs). Rather than becoming obsolete, the role of formal ontologies has evolved to become more essential as a "Grounding Layer." While LLMs excel at natural language understanding, these probabilistic models often lack the formal structural rigor required for authoritative enterprise environments. The comparative benchmark demonstrated this directly: the LLM-based extraction pipeline was identical for both ontologies, yet the quality and type of analytical output differed fundamentally based on the ontological structure. This confirms that the ontology—not the language model—determines the analytical ceiling of the system.

9.5. Limitations

This study has several limitations that should be acknowledged:

1. LLM dependency: The comparative benchmark relies on a single LLM (Qwen3.6-Plus) for entity extraction. Absolute entity counts may vary across different models or runs; we therefore emphasize relative distributions and comparative metrics rather than absolute figures.
2. Corpus scope: The real-world corpus consists of 13 unique Thai-language Dx documents. While heterogeneous in format, it is moderate in size and linguistically specific. Generalization to other languages and larger corpora requires further validation.
3. Absence of precision/recall: Ground truth could not be precisely enumerated because the corpus mixes project records with non-project content at differing granularities. We rely on independent cross-validation of distributions rather than precision/recall against a fixed denominator.
4. Scalability evaluation scope: The synthetic scalability experiments used RDFLib's in-memory store. Production deployments with dedicated triple stores (e.g., Apache Jena Fuseki, GraphDB) would likely exhibit different performance characteristics.
5. Ontology evolution: The current evaluation represents a snapshot of Dx-Onto's design. As the domain evolves, the ontology will require periodic updates, and the long-term maintenance burden has not been empirically assessed.

10. Future Roadmap: Synergizing Dx-Onto with Large Language Models

The evaluation in this study confirms that Dx-Onto provides a robust semantic backbone for querying digital transformation knowledge. However, we acknowledge a critical question in the current AI landscape: In the era of Large Language Models (LLMs) and Generative AI, is manually engineering an ontology still necessary? Our answer is a definitive yes. While LLMs excel at natural language understanding, they suffer from well-documented limitations in enterprise settings, specifically hallucinations and a lack of domain-specific context. An LLM alone cannot reliably query a fragmented repository of private organizational spreadsheets and guarantee 100% accuracy. Therefore, we position Dx-Onto not as a legacy artifact, but as the critical "Grounding Layer" (or Source of Truth) for a future Hybrid LLM-Ontology Architecture.

We explicitly note that the LLM integration described in this section constitutes future work and is not implemented in the current study. This paper focuses exclusively on the ontology design, scalability validation, and comparative benchmark evaluation. The following architectural overview outlines our planned direction.

We envision a system where the accessibility of LLMs is combined with the precision of Dx-Onto, functioning as a Retrieval-Augmented Generation (RAG) workflow. As illustrated in Figure 7, the architecture operates in three stages:

1. Ingestion: Heterogeneous documents (PDFs, images, spreadsheets) are ingested and mapped to the Dx-Onto structure via an automated "DX Indexer" pipeline (as partially demonstrated in the comparative benchmark).
2. Query Translation: When a user poses a natural language question, the LLM translates the intent into a precise SPARQL query targeting the Dx-Onto knowledge base. The ontology's class hierarchy and property definitions constrain the query space, preventing the LLM from generating queries against non-existent concepts.
3. Grounded Response: The SPARQL query returns verified, structured results from the ontology. These results are fed back to the LLM, which synthesizes a natural language response grounded in ontology-verified facts rather than probabilistic generation.

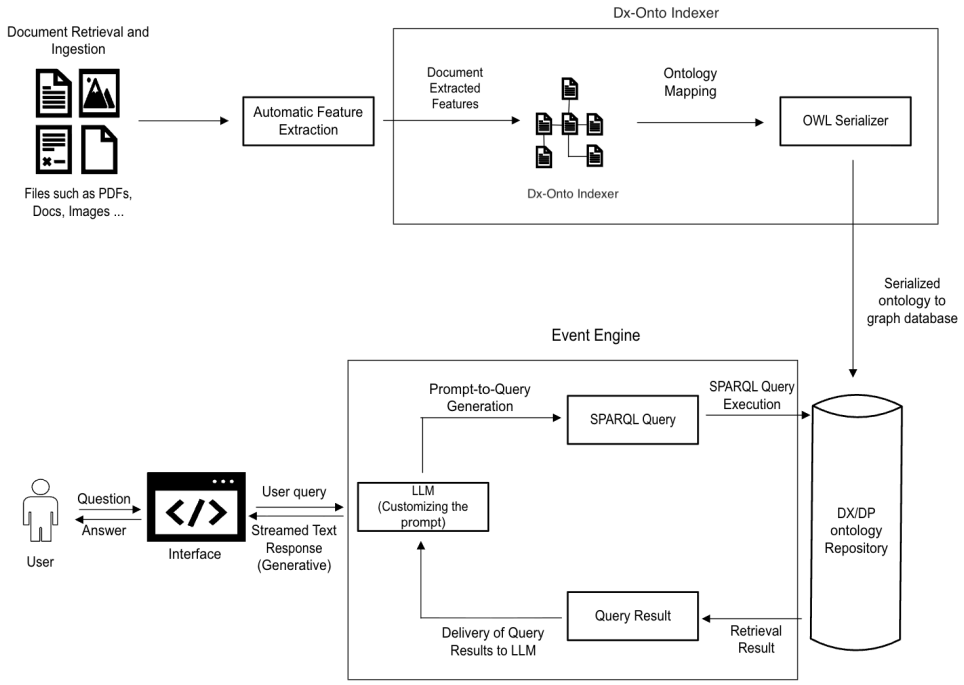


Figure 7. Envisioned Architecture: Using Dx-Onto as a verified knowledge base to ground LLM responses.

This hybrid approach addresses two fundamental challenges in modern knowledge management. First, it prioritizes reliability over probability; by forcing the AI to query the ontology, we prevent the "black box" hallucination problem common in pure LLM approaches. Second, it enables structured access to unstructured data; as demonstrated in this paper, Dx-Onto unifies diverse artifacts, allowing the future LLM to "read" and "reason" across diverse file types that are otherwise inaccessible to standard database queries.

11. Conclusions

This study addressed the critical challenge of knowledge fragmentation in Digital Transformation (Dx) initiatives by designing, developing, and validating Dx-Onto, a core domain ontology implemented in OWL for representing and managing Dx project knowledge. Unlike traditional document storage systems, Dx-Onto provides a unified semantic layer that connects heterogeneous project artifacts—from strategic dimensions to operational spreadsheets and multimedia files—into a machine-readable knowledge graph.

A two-pronged evaluation strategy was employed. First, scalability experiments using synthetic datasets confirmed that Dx-Onto maintains predictable query performance across volumes from 10 to 1000 projects, with complex multi-dimensional queries exhibiting particularly strong scalability. Second, a comparative benchmark against the Core Ontology

of Organisational Transformation (COOT) using a real-world corpus of 13 unique Thai Dx documents demonstrated that Dx-Onto achieves a higher domain fit (85.2% vs. 77.3%) and—critically—enables transformation-specific portfolio analytics (phase distributions, dimension analysis) that are structurally impossible under a general-purpose baseline. Both ontologies extracted comparable total entity volumes, confirming that the advantage of Dx-Onto lies in the analytical expressiveness of its domain-specific vocabulary rather than in raw extraction capacity.

Despite these contributions, we acknowledge several limitations: the dependence on a single LLM for the extraction pipeline, the moderate size and Thai-language specificity of the real-world corpus, the absence of precision/recall against a fixed ground truth, and the inherent challenges of ontology maintenance over time. These limitations motivate future work in three directions: (1) expanding the corpus to multilingual and cross-organizational datasets, (2) implementing the proposed Hybrid LLM-Ontology Architecture to bridge the usability gap between SPARQL expertise and natural language interaction, and (3) conducting longitudinal studies on ontology evolution and organizational adoption.

Author Contributions: S.B.-a.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing—original draft preparation, Visualization. C.A.: Conceptualization, Methodology, Resources, Writing—review and editing, Supervision, Project administration, All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: The authors gratefully acknowledge the domain insights and practical perspectives gathered from various public and private sector digital transformation initiatives, which informed the development of Dx-Onto. The authors also extend their appreciation to their academic supervisors for their guidance and constructive feedback, which greatly supported the design, implementation, and validation of the ontology throughout this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lei, B. Research on the Impact of Digital Transformation on Enterprise Operational Efficiency. *Front. Bus. Econ. Manag.* **2025**, *18*, 95–97. [CrossRef]
2. Aras, A.; Büyüközkan, G. Digital Transformation Journey Guidance: A Holistic Digital Maturity Model Based on a Systematic Literature Review. *Systems* **2023**, *11*, 213. [CrossRef]
3. Heisig, P.; Suraj, O.A.; Kianto, A.; Kemboi, C.; Perez Arrau, G.; Fathi Easa, N. Knowledge management and business performance: Global experts' views on future research needs. *J. Knowl. Manag.* **2016**, *20*, 1169–1198. [CrossRef]
4. Durst, S.; Edvardsson, I.R.; Foli, S. Knowledge management in SMEs: A follow-up literature review. *J. Knowl. Manag.* **2023**, *27*, 25–58. [CrossRef]
5. Zaoui, F.; Assoul, S.; Souissi, N. What Are the Main Dimensions of Digital Transformation? Case of an Industry. *Int. J. Recent Technol. Eng. (IJRTE)* **2019**, *8*, 9962–9970. [CrossRef]
6. Vial, G. Understanding digital transformation. In *Managing Digital Transformation*; Routledge: London, UK, 2021; pp. 13–66. [CrossRef]
7. Verhoef, P.C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Qi Dong, J.; Fabian, N.; Haenlein, M. Digital transformation: A multidisciplinary reflection and research agenda. *J. Bus. Res.* **2021**, *122*, 889–901. [CrossRef]
8. Company, M. Unlocking Success in Digital Transformations. 2018. Available online: https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/unlocking-success-in-digital-transformations (accessed on 29 October 2018).
9. Vial, G. Understanding digital transformation: A review and a research agenda. *J. Strateg. Inf. Syst.* **2019**, *28*, 118–144. [CrossRef]
10. Sehgal, A. Introduction to Digital Transformation. In *Demystifying Digital Transformation*; Apress: Berkeley, CA, USA, 2023; pp. 1–37. [CrossRef]

11. Yadav, G.N.S.; Seranmadevi, R. The Digital Transformation: Crafting Customer Engagement Strategies for Success. In *Digital Technologies, Ethics, and Decentralization in the Digital Era*; IGI Global: Hershey, PA, USA, 2024; pp. 80–98. [CrossRef]
12. Mulder, H.; Johnson, J.; Meijer, K.; Crear, J. Investigating 5, 140 Digital Transformation Projects. In *Digital Enterprises*; Springer Nature Switzerland: Cham, Switzerland, 2023; pp. 13–31. [CrossRef]
13. Deshmukh, R.R.; Raut, A.B. Ontology based Semantic Information Retrieval System using Data Ranking. In *Proceedings of the 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7. [CrossRef]
14. Kaltenegger, J.; Frandsen, K.M.; Petrova, E. An ontology-driven framework for digital transformation and performance assessment of building materials. *Build. Environ.* **2025**, *271*, 112565. [CrossRef]
15. Spoladore, D.; Pessot, E. An evaluation of agile Ontology Engineering Methodologies for the digital transformation of companies. *Comput. Ind.* **2022**, *140*, 103690. [CrossRef]
16. Bogea Gomes, S.; Santoro, F.M.; Mira da Silva, M.; Pinto, P.; Guizzardi, G. Towards a core ontology of organisational transformation. *Appl. Ontol.* **2023**, *18*, 31–70. [CrossRef]
17. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]
18. Guarino, N. Formal ontology and information systems. In *Proceedings of FOIS'98*; IOS Press: Trento, Italy, 1998; pp. 3–15.
19. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]
20. McGuinness, D.L.; Van Harmelen, F. OWL Web Ontology Language overview. In *W3C Recommendation*; W3C, Inc.: Wakefield, MA, USA, 2004. Available online: https://www.w3.org/TR/owl-features/ (accessed on accessed on 10 February 2026).
21. Henderson, J.C.; Venkatraman, N. Strategic alignment: Leveraging information technology for transforming organizations. *IBM Syst. J.* **1993**, *32*, 472–484. [CrossRef]
22. Tornatzky, L.G.; Fleischer, M. *The Processes of Technological Innovation*; Lexington Books: Lanham, MD, USA, 1990.
23. Yang, Q.; Aman, A.B.; Zaki, H.O.; Baharin, R. Digital trade transformation for SMEs in Asia Pacific using the technology-organization-environment framework: A systematic literature review. *Veredas Direito* **2026**, *23*, e234766. [CrossRef]
24. Teece, D.J.; Pisano, G.; Shuen, A. Dynamic capabilities and strategic management. *Strateg. Manag. J.* **1997**, *18*, 509–533. [CrossRef]
25. Zhou, J.; Kuang, M. Digital transformation, dynamic capabilities, and aggressive corporate strategy. *Int. Rev. Econ. Financ.* **2025**, *104*, 105905. [CrossRef]
26. Westerman, G.; Bonnet, D.; McAfee, A. *Leading Digital: Turning Technology into Business Transformation*; Harvard Business Press: Brighton, MA, USA, 2014.
27. Teichert, R. Digital transformation maturity: A systematic review of literature. *Acta Univ. Agric. Silvic. Mendel. Brun.* **2019**, *67*, 1673–1687. [CrossRef]
28. Warner, K.S.; Wäger, M. Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Plan.* **2019**, *52*, 326–349. [CrossRef]
29. Nonaka, I. A dynamic theory of organizational knowledge creation. *Organ. Sci.* **1994**, *5*, 14–37. [CrossRef]
30. Alavi, M.; Leidner, D.E. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Q.* **2001**, *25*, 107–136. [CrossRef]
31. Pappas, I.O.; Mikalef, P.; Giannakos, M.N.; Kourouthanassis, P.E.; Lekakos, G. Big data and business analytics ecosystems: Paving the way towards digital transformation and sustainable societies. *Inf. Syst. E-Bus. Manag.* **2018**, *16*, 479–491. [CrossRef]
32. MIT Sloan Management Review. Strategy Drives Digital Transformation. 2020. Available online: https://sloanreview.mit.edu/projects/strategy-drives-digital-transformation/ (accessed on 4 February 2026).
33. Corcho, O.; Fernández-López, M.; Gómez-Pérez, A. Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data Knowl. Eng.* **2003**, *46*, 41–64. [CrossRef]
34. Simperl, E. Reusing ontologies on the Semantic Web: A feasibility study. *Data Knowl. Eng.* **2009**, *68*, 905–925. [CrossRef]
35. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From ontological art towards ontological engineering. In Proceedings of the AAAI-97 Spring Symposium Series on Ontological Engineering, Palo Alto, CA, USA, 24–25 March 1997; pp. 33–40.
36. Noy, N.F.; McGuinness, D.L. *Ontology Development 101: A Guide to Creating Your First Ontology*; Stanford Knowledge Systems Laboratory Technical Report KSL-01-05; Stanford Knowledge Systems Laboratory: Palo Alto, CA, USA, 2001.
37. Deliyska, B.; Ivanova, A. Ontological Metamodel of Sustainable Development. In *Encyclopedia of Data Science and Machine Learning*; IGI Global Scientific Publishing: Hershey, PA, USA, 2023; pp. 2890–2904. [CrossRef]
38. Halper, M.; Soldatova, L.N.; Brochhausen, M.; Sabiu Maikore, F.; Ochs, C.; Perl, Y. Guidelines for the reuse of ontology content. *Appl. Ontol.* **2023**, *18*, 5–29. [CrossRef]
39. Kebede, R.; Moscati, A.; Tan, H.; Johansson, P. Integration of manufacturers' product data in BIM platforms using semantic web technologies. *Autom. Constr.* **2022**, *144*, 104630. [CrossRef]
40. Dermeval, D.; Albuquerque, J.; Bittencourt, I.I.; Isotani, S.; Silva, A.P.; Vassileva, J. GaTO: An Ontological Model to Apply Gamification in Intelligent Tutoring Systems. *Front. Artif. Intell.* **2019**, *2*, 13. [CrossRef] [PubMed]

41. Cantador, I.; Bellogín, A.; Castells, P. A multilayer ontology-based hybrid recommendation model. *AI Commun.* **2008**, *21*, 203–210. [CrossRef]
42. Luschi, A.; Petraccone, C.; Fico, G.; Pecchia, L.; Iadanza, E. Semantic Ontologies for Complex Healthcare Structures: A Scoping Review. *IEEE Access* **2023**, *11*, 19228–19246. [CrossRef]
43. Sampath Kumar, V.R.; Khamis, A.; Fiorini, S.; Carbonera, J.L.; Olivares Alarcos, A.; Habib, M.; Goncalves, P.; Li, H.; Olszewska, J.I. Ontologies for Industry 4.0. *Knowl. Eng. Rev.* **2019**, *34*, e17. [CrossRef]
44. Witt, A.; Huggins, A.; Governatori, G.; Buckley, J. Converting copyright legislation into machine-executable code: Interpretation, coding validation and legal alignment. In *Proceedings of the ICAIL '21: Eighteenth International Conference on Artificial Intelligence and Law*; ACM: New York, NY, USA, 2021; pp. 139–148. [CrossRef]
45. Olsina, L.; Becker, P.; Papa, M.F. The Project Management Ontology called ProjectCO: Architectural Aspects, Concepts, and Usefulness. In *Proceedings of the Anais do XXVII Congresso Ibero-Americano em Engenharia de Software (CIbSE 2024)*; Sociedade Brasileira de Computação: Porto Alegre, Brazil, 2024; pp. 61–75. [CrossRef]
46. Bastos, E.C.; Barcellos, M.P.; de Almeida Falbo, R. Using Semantic Documentation to Support Software Project Management. *J. Data Semant.* **2018**, *7*, 107–132. [CrossRef]
47. Zaouga, W.; Rabai, L.B.A. A Smart Ontology for Project Risk Management Based on PMI's Framework. In *Proceedings of Fifth International Congress on Information and Communication Technology*; Springer: Singapore, 2020; pp. 218–228. [CrossRef]
48. Moudoubah, L. Development of Ontology for Semantic Structure of Strategic Alignment Framework for IT Projects Combining PMBOK, PMI and VAL IT. *Int. J. Intell. Eng. Syst.* **2021**, *14*, 458–468. [CrossRef]
49. Tkachenko, K.; Tkachenko, O.; Tkachenko, O. Modeling and using intelligent multi-agent system in smart city: Ontological approach. *Transp. Syst. Technol.* **2023**, *42*, 45–57. [CrossRef]
50. Wintercorn, O.; Van Deventer, J.; Paniagua, C. Unlocking the Power of Digital Transformation: The Role of Ontologies. In *Proceedings of the 2025 IEEE International systems Conference (SysCon)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [CrossRef]
51. Chan, H.K.H.; Long, C.; Yan, D.; Wong, R.C.W.; Lu, H. Fraction-Score: A Generalized Support Measure for Weighted and Maximal Co-location Pattern Mining. *IEEE Trans. Knowl. Data Eng.* **2023**, *36*, 1582–1596. [CrossRef]
52. Putrama, I.M.; Martinek, P. Heterogeneous data integration: Challenges and opportunities. *Data Brief* **2024**, *56*, 110853. [CrossRef] [PubMed]
53. Borowicc, S.; Alves-Souza, S. Heterogeneous Data Integration: A Literature Scope Review. In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024)*; SciTePress: Setúbal, Portugal, 2024; Volume 1, pp. 189–200. [CrossRef]
54. Ranatunga, S.; Ødegård, R.S.; Jetlund, K.; Onstein, E. Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 52. [CrossRef]
55. Kone, N.; Mahesh, G. An ontology-driven bi-directional workflow for integrating project management data into the IFC standard. *J. Inf. Technol. Constr. (ITcon)* **2025**, *30*, 1770–1790. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.