Managing Data Heterogeneity for Ontology-Driven Models: Application to Gamified E-Learning Contexts

Yara Gomaa[1,2][a], Christine Lahoud[3][b], Marie-Hélène Abel[1][c] and Sherin Moussa[2,4][d]

[1]*Laboratoire HEUDIASYC, Université de Technologie de Compiègne, Sorbonne Universités, 57 Avenue Landshut, Compiègne 60200, France*

[2]*Laboratoire Interdisciplinaire de l'Université Française d'Egypte (UFEID Lab), Université Française d'Egypte, 21 Ismailia Desert Road Shorouk City, Cairo 11837, Egypt*

[3]*CIAD UR 7533,Univ. BourgogneFranche-Comté, UTBM, F-90010 Belfort, France*

[4]*Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt*

Keywords: Gamification, E-Learning, Ontology, Data Collection, Data Integration.

Abstract: Data heterogeneity within gamified e-learning systems exposes a challenge for ontology-driven models, specifically ontology-based recommender systems. These systems can help teachers who are unfamiliar with gamification by offering personalized recommendations to gamify their pedagogical resources. Yet, developing such systems requires collecting and integrating diverse data about users, resources, and game elements, originating from multiple sources, like learning management systems and educational repositories, each with varying formats and inconsistent semantics. This paper proposes an approach to manage the complexities of collecting and preparing heterogeneous data for an ontology-driven model within gamified e-learning contexts. A full overview is provided on the data workflow, which consists of two main phases: (1) Data collection, which combines automated techniques through APIs and web scraping, and (2) Data Integration by means of mapping the collected data into our Teacher in Gamified e-learning Context (TGC) ontology to produce coherent and semantically enriched structure. The resulting data repository facilitates semantic queries, inference, and knowledge enrichment, overcoming challenges like cold-start scenarios and supporting the dynamic generation of personalized recommendations. This proposed approach aims to establish a robust approach that addresses the challenges of data heterogeneity, ensuring consistent and meaningful integration for ontology-based recommender systems in gamified e-learning contexts.

1 INTRODUCTION

Gamified E-learning systems have revolutionized education by enhancing engagement and fostering interactive learning experiences (Bennani et al., 2024; Jumaa et al., 2017). Gamification, in this regard, introduces game-like experiences to e-learning environments through tools commonly known as game elements . These game elements include points, badges, leaderboards, etc (Maher et al., 2020). Despite their potential, many teachers face challenges in understanding and implementing gamification effectively. This unfamiliarity hinders the ability to optimize gamification to serve the purpose behind it (Gomaa et al., 2024). Addressing this gap requires systems that can assist teachers to link their objectives with the relevant gamification strategy; by means of recommending relevant game elements or gamified resources. Ontology-driven models offer a promising solution to this problem. By semantically structuring data, such systems can provide a robust framework for generating personalized suggestions tailored to the pedagogical and gamification needs of teachers (Bakhouyi et al., 2019). In addition, ontology-based recommender systems can mitigate the cold-start

[a] https://orcid.org/0000-0001-9853-2416
[b] https://orcid.org/0000-0002-4520-634X
[c] https://orcid.org/0000-0003-1812-6763
[d] https://orcid.org/0000-0001-9593-6909

Gomaa, Y., Lahoud, C., Abel, M.-H. and Moussa, S.
Managing Data Heterogeneity for Ontology-Driven Models: Application to Gamified E-Learning Contexts.
DOI: 10.5220/0013496700003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 1*, pages 726-736
ISBN: 978-989-758-746-7; ISSN: 2184-5026

problem by inferring recommendations through established relationships, even with limited initial data (Nymfodora-Maria Raftopoulou et al. 2023).

However, preparing data for such systems can be challenging due to the heterogeneity of data. An ontology-based recommender system expects collecting data about three main categories, namely (1) users, whether a teacher or learner, (2) pedagogical resources, and (3) game elements. Each category deals with further detailed data, for instance, teachers are identified by personal information, demographic data, institutional data, pedagogical and gamification-related data. Such data can be found in various sources, including educational repositories, learning management systems, plugin libraries, and user-generated sources, i.e., questionnaires (Joy et al., 2021).

Preparing these diverse data sources for an ontology-driven model demands a comprehensive data collection and integration strategy:

- Data Collection: employing automated and semi-automated techniques to gather the required data. Automated methods rely heavily on APIs, which provide structured access to data from platforms, like Moodle[5]. These APIs allow for real-time or scheduled retrieval of specific datasets. Another technique is web scraping, which extracts the HTML webpage content (Mansouri et al., 2022). On the other hand, semi-automated methods often include gathering qualitative data through questionnaires. This method is helpful when human input is needed, for example, for teachers identifying their gamification experiences (Joy et al., 2021).

- Data Integration: once collected, the heterogeneous data must be aligned to a shared ontology, which involves converting raw inputs into Resource Description Framework (RDF) triplets and mapping them onto the domain concepts relevant to gamification and pedagogy. Consistency checks and cleaning processes are essential to reconcile conflicting data formats and semantic inconsistencies (Villegas-Ch et al. 2023).

Numerous studies focused on data heterogeneity in e-learning systems. However, limited focus was given to addressing data heterogeneity in ontology-driven models in gamified context. This leads to the main research question of our study:

How to collect and integrate heterogeneous data from multiple sources for ontology-based models in gamified e-learning contexts?

To address this question, we propose an approach that systematically aggregates and integrates diverse data into an ontology-driven model. Our methodology focuses on collecting comprehensive data from multiple sources, including educational content, teacher profiles, and gamification elements. By emphasizing the teacher's role, we aim to make them more familiar with applying gamification techniques to their teaching practices. Additionally, our approach allows for the dynamic adaptation of game elements, meaning that educators can remove or add specific game elements as needed to optimize the learning experience. This flexibility ensures that the full potential of gamification can be utilized, enhancing both the educational process and the learner's engagement. The rest of the paper is structured as follows: Section 2 explores the main studies related to our work, involving data collection and integration techniques. Followed by section 3 for our proposed data collection and data integration approach. Section 4 discusses the proposed work. Then, section 5 presents the conclusion of our study, highlighting the future directions to adopt.

2 RELATED WORK

This section explores the various data collection techniques employed in the main studies within the e-learning and gamified e-learning domains, focusing on the types of data collected, the methods used for their integration, and the role of ontologies in structuring such data.

2.1 Data Management in E-Learning Systems

This study focused on collecting and integrating heterogeneous big data from E-Learning systems, particularly Moodle LMS for analytics (Otoo-Arthur & van Zyl, 2020). The integration process did not involve ontology but relied on a big data framework (biDeL) using Apache Spark, Hadoop and Flume for data processing. The main goal was to enhance learning analytics, enabling institutions to gain insights into learner behavior and instructional effectiveness. Data was collected through log tracking, system interactions, and machine learning. Challenges included data privacy, governance, and

[5] Moodle

the complexity of handling high volume, high velocity and varied data sources, necessitating advanced frameworks for scalability and security.

For enhancing the learning analytics within an e-learning systems such as MOOCs, data collected was mainly about the learners' interactions, profiles, and engagement within the e-learning systems. These data were collected from Experience API (xAPI) and IEEE Standard for Learning Technology (SLT) (Del Blanco et al., 2013).

The previous studies focused on the non-gamified context, limiting learner preferences, datatypes collected, and enhanced learning outcomes. Unlike these studies, where collected data was primarily used for learning and analytics, our approach focuses on structuring data within an ontology, enabling continuous refinement, semantic enrichment, and enhanced knowledge representation.

2.2 Data Management in Ontology-Based E-Learning Systems

This paper presents a semantic web-based approach to enhance E-Learning systems interoperability using RDF and next generation SCORM specifications (Bakhouyi et al., 2019). It addresses interoperability challenges in LMS platforms like Moodle by transforming JSON data into RDF with JSON-LD, ensuring seamless data sharing between xAPI and LMS. The study highlights Semantic Web Technologies' potential to improve content exchange, tracking, and analytics, enabling better accessibility, learning personalization, and integration of mobile applications into e-Learning ecosystem.

The study proposes a hybrid Adaptive Educational e-Learning System (AEeLS) integrating artificial intelligence and semantic web technologies to personalize learning (Demertzi & Demertzis, 2020). AEeLS adapts educational content based on student skills and experience by using ontology matching and a recommendation system. The ontology matching employs semi-supervised machine learning to align educational resources, while the recommendation mechanism uses collaborative and content-based filtering to suggest relevant learning materials. The system improves interoperability, efficiency, and adaptability in e-learning. Experiments on datasets from OAEI 2014, ADRIADNE, and MERLOT demonstrate their super superior performance in matching and recommending educational content.

Resources were considered in more detail in (Bouihi & Bahaj, 2019), focusing on data collection and integration related to learning content, learner profiles, social interactions, and learner activities to enhance personalized learning through a semantic web-based recommendation system. It used learning objects modelled with standards like LOM and SCORM as the core data, enriched with contextual information like learner's history, performance, and social connections, and represented using ontologies like FOAF. Accordingly, these data were utilized to personalize recommendations for learners. However, the study could have benefited from extending FOAF ontology to further model the various parameters of a resource and a learner.

The challenges of data interoperability in collaborative e-learning systems were tackled in (Masud, 2016), focusing on how to manage and share learning content across different systems with varying schemas. The data collected included learning content and its metadata and shared between independent e-learning systems using schema-level and data-level mappings. Through query processing, relevant resources were retrieved based on user queries. This study enforced collaboration and sharing of resources between learners and teachers, as the resources were semantically categorized and managed, ensuring interoperability.

However, teachers-related data was explored in a limited context. This was overcome in (Nashed et al., 2022), in which an ontology-based approach was introduced to integrate diverse educational data related to teacher's personal, organizational, and contextual environments. It collected data about educational resources and mapped to Teacher-Context Ontology (TCO) using D2RQ mapping, which offered a comprehensive view of the teacher's professional and personal context. The approach used automated mapping techniques to merge data from various sources, facilitating its use in personalized educational systems, like recommender systems. Nevertheless, gamification in the e-learning context was not handled, where recommendations were fixed to the e-learning aspect only.

2.3 Data Management Gamified Context

Data heterogeneity in gamified e-learning systems was the focus of various studies. A gamified educational system was presented in (Nymfodora-Maria Raftopoulou et al. 2023) to enhance the gamified learning experience. It collected data, such as user experience rating, exercise completion times, success rates, and correct/incorrect answers, and used machine learning algorithms to analyze data, identify patterns in learner behaviour, and provide personalized feedback. The system used collaborative filtering and content-based filtering to recommend future exercises that meet each learner's strengths and weaknesses.

The authors focused on collecting and integrating student interaction data within a gamified e-learning system designed to develop computational thinking (Villalba-Condori et al., 2022). The system employed supervised learning techniques, specifically Support Vector Machines (SVM), to classify learning progress and recommend personalized content. The data came from students' in-game activities tracking performance, engagement, and decision making. While the study did not explicitly integrate an ontology-based approach, it leveraged machine learning and statistical models for recommendations and performance analytics. Challenges the complexity of accurately modeling student learning behaviors in a gamified environment.

Another adaptive gamification recommendation system was introduced in (Bennani et al., 2024), focusing on game elements adaptation to improve interactive learning environments. It collected both static and dynamic data about learners, such as their profile information, interaction, and performance metrics. Player type tests were used to assess the learner's preferred game elements. The integration process involves a matrix factorization process, in which the data is then utilized in collaborative filtering recommender system.

In (Alonso-Fernández et al., 2020), data was collected to predict learner's knowledge outcomes based on their interactions within a serious game. Data included personal information, like age and gender, knowledge data through pre and post-tests, and game interaction data, which were tracked through Experience API, such as scores, levels completed, etc.

Considering the presented studies, despite the progress in data integration within e-learning systems, a critical gap remains in addressing the heterogeneity of data across game elements, teachers, and the broader e-learning context. Most of the presented studies have primarily focused on learners and educational resources, with minor attention given to the teacher's profile or the varying data related to game elements. In particular, the complexities introduced by the teacher's context and gamification aspects have not been thoroughly investigated. This oversight highlights a significant opportunity to develop comprehensive systems that adopt teacher-specific data and game elements, enabling a holistic approach to a personalized and gamified e-learning experience.

3 PROPOSED APPROACHES

To prepare data for an ontology-driven model, it undergoes two main processes: (1) data collection and (2) data integration via ontology mapping. In this section, we introduce our proposed approach for the data collection process, the data sources, and the formats used to gather and store data.

3.1 Data Management Process

As the main goal of this study is to harmonize heterogeneous data within an ontology-driven model in gamified e-learning contexts, this section highlights the data workflow within a recommender system, along with its categories and data sources. Figure 1 provides a high-level overview of the entire data management process, including data collection, integration and their utilization within ontology-driven models such as ontology-based recommender system in gamified e-learning contexts. The process begins with collecting data about users (teachers and learners), resources, and game elements. For the users, data is broadly categorized into personal information, pedagogical and gamification-related data, and activity log. Users' personal information includes their name, age, gender, registered educational institution, and preferred language. Regarding the teacher's pedagogical-related information, data is related to the teaching style and pedagogical experience, to determine how skillful the teacher is in applying teaching strategies. In addition to the learning objectives, which define the teacher's intention behind a particular resource, in terms of the expected learning outcomes. As for the gamification-related data for teachers, it includes gamification experience, which reflects their ability to effectively utilize gamification techniques. The teacher's behavioral objectives refer to the targeted behavior expected from the learner towards a presented resource and playing style. On the other hand, the learner's pedagogical-related data refers to the learner's learning style, reflecting the learner's preferred way to receive and process the presented information. The learner's gamification-related data includes their player type, which helps identify the most effective methods to maintain their engagement. In addition, the activity log refers to the teacher's and learner's interactions throughout the system, including their feedback, ratings, and behavior.

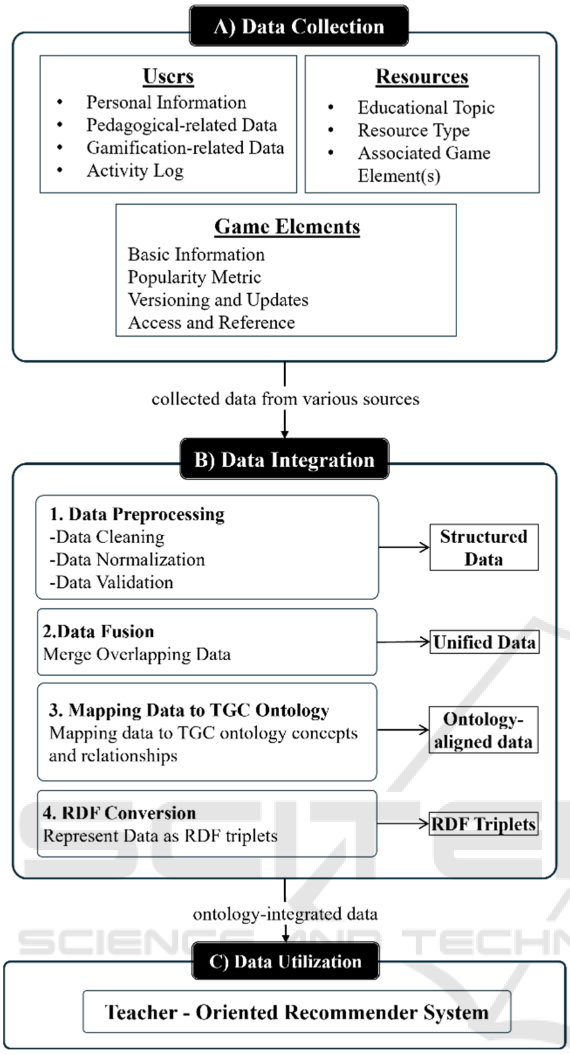

Figure 1: Data Management for Ontology-driven Model.

Concerning resources, the required data mainly includes the related educational topic, the resource's type (video, word document, audio file, etc.), in addition to the game elements associated with that resource (if any). The available game elements need to be identified with basic information, like title, description of its functionality, along with popularity metrics, like total downloads and likes, etc. The game element's version indicates whether it is up-to-date or obsolete, and finally, a reference to the game element, such as URL, can be provided as well. After gathering the needed data, the next step is to pre-process and clean data to ensure consistency with other data sources. This involves standardizing the format of the collected data and handling any missing or incomplete information. Once cleaned, data is mapped to the ontology to ensure it could be effectively queried and utilized in a recommendation process. Having the data integrated into our ontological model, it is utilized by recommender systems to provide personalized recommendations for teachers. For the following sections, we adopt a unified scenario as an example of how the data are collected and integrated for recommender systems. This scenario is illustrated in two partitions as scenario-part 1 and scenario-part 2 in sections 3.1 and 3.2 respectively.

Scenario Overview: *Mr. Robert Sedgewick is a teacher who is unfamiliar with gamification but wishes to integrate game elements into his course on "Computer Science Programming". To demonstrate the data collection and integration processes, an existing Coursera course entitled as "Computer Science: Programming with a Purpose" is adopted as a case study. This course has predefined data, such as the course modules and resources. Moreover, for the gamification strategies, game element plugins are adopted from Moodle. This demonstrates how the relevant data are collected and integrated to be ready for utilization by ontology-based recommender systems.*

3.2 Data Collection

The data collection process involves extracting and aggregating data from several key categories. Referring to our scenario to demonstrate the data collection process:

Scenario-part1: "*To provide Mr. Robert Sedgewick with personalized recommendations on the gamification of his course, his personal data, pedagogical-related data, gamification-related data, and activity log are collected. In addition, the targeted audience knowledge level, and the learning objectives are collected, as well as the course's structure itself along with its modules and resources from Coursera.".*

Table 1 presents the typical data sources for each category, along with their expected data formats. Educational institutions include universities, schools, etc., whereas educational platforms are e-learning management systems, such as Moodle. Users refer mainly to teachers and learners, where various types of data are required from them, scattered across various data sources. Typical personal information can be collected from their educational institutions or from their user profile on the educational platforms. However, there are other data related to pedagogy and gamification that can be collected through questionnaires, as they are particularly difficult to find explicitly through educational platforms or institutions. Such data includes teacher's teaching style, teacher's playing style, learner's learning style, and learner's player type. These data can be initially identified through questionnaires like the Fedler Silverman Learning Style Model FSLM (Felder & Spurlin, 2005), or Marczewski's player type model (Tondello et al., 2016). Furthermore, activity logs refer to the recorded user's interaction activity, determining the user's behavior, feedback, accessed resources, ratings, etc. However, the user's interactivity within an educational platform can only be accessed and collected through that platform. Furthermore, Resources are shared through digital form, which can be accessible through educational platforms or institutions.

Table 1 Data types and their corresponding sources for ontology integration.

Data Type		Data Source	Data Format
User	Personal Information	Educational Institutions, Educational Platforms	CSV, JSON
	Pedagogy-related Data	Online questionnaire	CSV, JSON, database
	Gamification-related Data	Online questionnaire	CSV, JSON, database
	Activity Log	Educational Platforms	CSV, JSON, TXT, XML
Resources	Educational Institutions, Educational Platforms	Educational Institutions, Educational Platforms	CSV, JSON
Game Elements	Educational Platforms	Educational Platforms	XML

Data are gathered about a programming course on Coursera[6] that was done semi-automatically through web scrapping by BeautifulSoup API. As shown in the resulted JSON file in Figure 2, it contains basic information about the course, including course title, course instructors, the course language, the targeted level of audience, skills associated with the course, and the course modules. Each module has title and a set of resources, where each resource is identified by its type and duration. On the other hand, game elements are expected to be associated to a resource. Game elements are plugins that can be shared across educational platforms. In this regard, Moodle gamification plugins are used as a case study for data collection of game elements. Figure 3. shows a screenshot from Moodle demonstrating part of the list of game elements in the platform, with detailed description for each one.

```
"Course Title": "Computer Science: Programming with a Purpose",
"Instructors": [
    "Robert Sedgewick",
    "Kevin Wayne"
],
"Level of Audience": "Beginner level",
"Skills": [
    "Programming Principles",
    "Computer Science",
    "Algorithms",
    "Java Programming"
],
"Taught Language": "Taught in English",
"Available Languages": "22 languages available",
"University": "Princeton University",
"Modules": [
    {
        "Module Name": "CONDITIONALS AND LOOPS",
        "Activities": [
            {
                "Activity Name": "Conditionals: the if statement",
                "Type": "Video",
                "Duration": "6 minutes"
            },
            {
                "Activity Name": "Loops: the while statement",
                "Type": "Video",
                "Duration": "11 minutes"
            },
```


Figure 2: Part of JSON file for the course, teachers, lessons, and resources through BeautifulSoup.

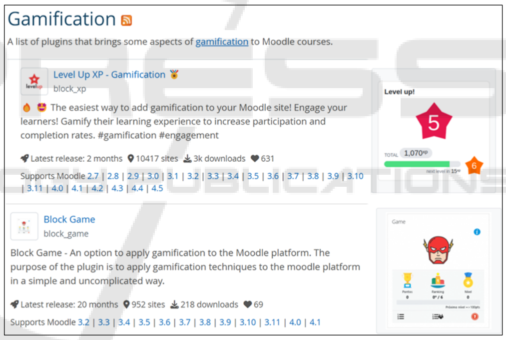

Figure 3: Screenshot from Moodle for game elements plugins.

This data were then collected in JSON file as shown in Figure 4, making it ready to be integrated into our ontology.

3.3 Data Integration

To harmonize the data gathered from the data collection phase, data is mapped to our Teachers in Gamified e-learning Context (TGC) Ontology. The following scenario demonstrates the work done in the data integration process:

<u>Scenario-part 2</u>: *"The collected data is then integrated to our TGC ontology, where the structured*

[6] Coursera

data of Mr. Robert, the game elements from Moodle, and the course are mapped to our TGC ontology based on the defined rules. The resultant ontology-integrated data is in RDF format."

TGC ontology focuses on the teacher's perspective within the gamified e-learning context. It reuses the existing ontologies MC2 and TCO (Abel et al., 2007; Nashed et al., 2021). MC2 ontology is concerned with the collaboration and shareability between resources, while TCO is concerned with defining the teacher in several contexts, including working and living environments. Our proposed TGC ontology utilizes MC2 and TCO to have a comprehensive semantic structure for the gamification aspect, together with the teachers and pedagogical resources. The automated technique used for mapping the collected data to our TGC ontology is carried out through RDFLib, a python library for working with RDF data. The mapping is performed by defining specific rules as shown in (1) and (2):

Raw Data → *Instructors{Robert Sedgewick, Kevin Wayne}* (1)

Raw Data → *Course Title: "Computer Science: Programming with a Purpose"* (2)

Figure 5 shows that all attributes with the name *Instructors* are mapped to *TCO:Teacher*, where the values of *Instructors* are of type xsd:String, mapped to the data property *TCO:Teacher.has_name*. Similarly, *Course Title* is mapped to *TCO:Course.has_title*, where the object property *TCO:teaches* associates the *TCO:Teacher* with *TCO:Course*, as shown in (3):

RDF Triplet → *Robert Sedgewick, Teaches, Computer Science: Programming with a Purpose* (3)


```json
    {
        "title": "Level Up XP - Gamification 🎖",
        "link": "https://moodle.org/plugins/block_xp",
        "published": "Sun, 10 Nov 2024 12:24:38 GMT",
        "summary": "by Frédéric Massart ★.  <p>💧 🤩 The easiest way
to add gamification to your Moodle site! Engage your learners! Gamify their
learning experience to increase participation and completion rates.
#gamification #engagement</p>"
    },
    {
        "title": "Block Game",
        "link": "https://moodle.org/plugins/block_game",
        "published": "Wed, 31 May 2023 12:55:01 GMT",
        "summary": "by Jota Costa.  <p>Block Game - An option to apply
gamification to the Moodle platform.  The purpose of the plugin is to apply
gamification techniques to the moodle platform in a simple and
uncomplicated way.</p>"
    },
```


Figure 4: Part of JSON file for the game elements through BeautifulSoup.


```xml
<rdf:Description rdf:about="http://www.hds.utc.fr/tco/tbox#Kevin_Wayne">
  <rdf:type rdf:resource="http://www.hds.utc.fr/tco/tbox#Teacher"/>
  <tco:has_name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Kevin Wayne</tco:has_name>
  <tco:teaches rdf:resource="http://www.hds.utc.fr/tg/tbox#Computer_Science_Programming_with_a_Purpose"/>
  <tco:works_at rdf:resource="http://www.hds.utc.fr/tco/tbox#Princeton_University"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.hds.utc.fr/tco/tbox#Robert_Sedgewick">
  <rdf:type rdf:resource="http://www.hds.utc.fr/tco/tbox#Teacher"/>
  <tco:has_name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Robert Sedgewick</tco:has_name>
  <tco:teaches rdf:resource="http://www.hds.utc.fr/tg/tbox#Computer_Science_Programming_with_a_Purpose"/>
  <tco:works_at rdf:resource="http://www.hds.utc.fr/tco/tbox#Princeton_University"/>
</rdf:Description>

<rdf:Description rdf:about="http://www.hds.utc.fr/tg/tbox#Computer_Science_Programming_with_a_Purpose">
  <rdf:type rdf:resource="http://www.hds.utc.fr/tco/tbox#Course"/>
  <tco:has_title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Computer Science: Programming with a Purpose</
  <tg:has_taught_language rdf:datatype="http://www.w3.org/2001/XMLSchema#string">English</tg:has_taught_language>
  <tg:has_university rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Princeton University</tg:has_university>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#CONDITIONALS_AND_LOOPS"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#ARRAYS"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#INPUT_AND_OUTPUT"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#FUNCTIONS_AND_LIBRARIES"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#RECURSION"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#PERFORMANCE"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#ABSTRACT_DATA_TYPES"/>
  <tco:consists_of rdf:resource="http://www.hds.utc.fr/tg/tbox#CREATING_DATA_TYPES"/>
</rdf:Description>
```


Figure 5: Teachers mapped automatically to RDF through RDFLib.

Moving to the course's modules and resources, Figure 6 shows the module mapped to TCO:Lesson, whereas its activities mapped to "TCO:Resource", associated together through the object property "*TCO:contain_Resource*". On the other hand, game elements collected were mapped to our TGC ontology. To understand the ontological structure of a game element, Figure 7. shows the game element class with its associated data properties. "*TGC:GameElementResource*" class has several performance metrics, like total downloads, total likes, total sites using this plugin, the plugin's latest release, and the latest version. These data are particularly helpful for a recommender system to recommend an appropriate game element that is relevant, popular, and up to date. The mapping of the game elements is performed by defining rules that associate each value attribute from the JSON file to a TGC:GameElementResource as shown in (4) and (5):

Raw Data → *title: "Block Game"* (4)

Raw Data → *published: "Wed,31 May 2023 12:55:01 GMT"* (5)

All attributes with the name title are considered as xsd:String and are mapped to the data property TGC:GameElementResource.hasTitle. For the attribute named published, it displays the date of publishing the plugin, however, our TGC ontology requires the duration in months since the latest published date. Therefore, it is pre-processed before the ontology mapping process. Figure 8 shows the resulted RDF file, where it illustrates that the game element has title "stash" and it is a named individual of type "TGC:GameElementResource" with all fetched data properties includes the resource title, its release date, a short description about this resource, and the resource's plugin Uniform Resource Identifier (URI) as shown in (6):

RDF Triplet → Block_Game, hasLatestRelease, "20"^^xsd:nonNegativeInteger (6)


```
<rdf:Description rdf:about="http://www.hds.utc.fr/tg/tbox#CONDITIONALS_AND_LOOPS">
  <rdf:type rdf:resource="http://www.hds.utc.fr/tco/tbox#Lesson"/>
  <tco:has_name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">CONDITIONALS AND LOOPS</tco:has_name>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Conditionals_the_if_statement"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Loops_the_while_statement"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#An_alternative_the_for_loop"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Nesting"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Debugging"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Supplements_for_Lecture_2"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Optional_Enrichment_on_Conditionals_and_Loops"/>
  <tco:contain_resource rdf:resource="http://www.hds.utc.fr/tg/tbox#Conditionals_and_Loops"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.hds.utc.fr/tg/tbox#RFK_L3_R5">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
  <rdf:type rdf:resource="http://www.hds.utc.fr/mc2sois/tbox#ReferenceKey"/>
  <tbox:indexed_by rdf:resource="http://www.hds.utc.fr/tg/tbox#TaskJavaClassesMethod"/>
```


Figure 6: Resources mapped automatically to RDF through RDFLib.

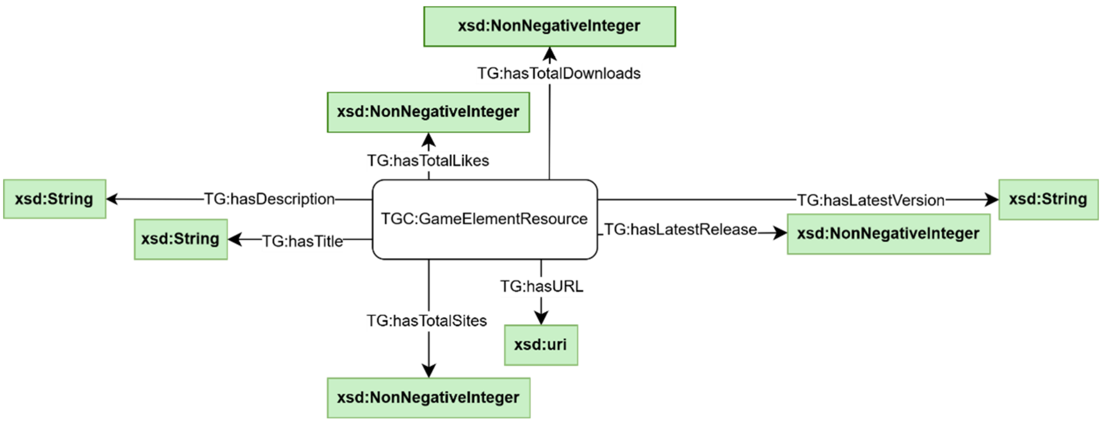

Figure 7: GameElementResource ontology class with its data properties.

```xml
<rdf:Description rdf:about="http://www.hds.utc.fr/tg/tbox#Stash">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
  <rdf:type rdf:resource="http://www.hds.utc.fr/tg/tbox#GameElementResource"/>
  <rdf:type rdf:resource="http://www.hds.utc.fr/tg/tbox#Reward"/>
  <tbox2:has_reference rdf:resource="http://www.hds.utc.fr/tg/tbox#RF_GE_STash"/>
  <tg:has_URL rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">https://moodle.org/plugins/block_stash</tg:has_URL>
  <tg:has_description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">by Adrian Greeve,
  Frédéric Massart ⭐. &amp;nbsp;&lt;p&gt;Add an inventory of items to your course and
   let your students find items by exploring the activities. #game #gamification&lt;/p&gt;</tg:has_description>
  <tg:has_latestRelease rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">3</tg:has_latestRelease>
  <tg:has_latestVersion>2.1.1</tg:has_latestVersion>
  <tg:has_title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Stash</tg:has_title>
  <tg:has_totalDownloads rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">1227</tg:has_totalDownloads>
  <tg:has_totalLikes rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">2003</tg:has_totalLikes>
  <tg:has_totalSites rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">736</tg:has_totalSites>
```


Figure 8: Game Elements mapped automatically to RDF through RDFLib.

4 DISCUSSION

In this research two main data sources are utilized to formalize the teacher's perspective within the given e-learning contexts.

The data related to the teacher and educational content including core structure lessons and resources were sourced from Coursera. Meanwhile, the game elements were treated as independent standalone plugins adopted from Moodle. This data was successfully structured and integrated within our proposed TGC ontology. Providing them with semantic meaning and associating them under a coherent unified structure.

Previous studies have aimed to structure data primarily. Focusing on learners with some studies incorporating aspects of course structure, and others focusing more on resource details. Other studies have explored game elements in combination with learner data leaving out the teacher's context and course data. This created a gap in linking these various elements. This paper takes an initial step toward bridging this gap by combining detailed data on users (teachers and learners), courses, and game elements into a single unified structure.

This study has encountered several challenges, firstly the limited scope of the game elements presents a constraint. These elements are typically integrated within e-learning systems. Standalone game elements are scarce, with limited variety and accessible through limited data sources. This narrow range may potentially limit the richness and adaptability of recommendations. In addition to this, Moodle platform introduced another challenge, as retrieving data automatically proved difficult. We attempted Moodle APIs, but this approach was not successful as the access was not authorized. The only accessible data was through partial web scraping, since not all properties of the game elements could be retrieved. Moreover, key teacher-related properties such as player type, teaching style, and gamification experience, are not typically included in teacher profiles in educational platforms or institutions. As a result, this data could not be retrieved automatically. Instead, questionnaires were necessary to obtain initial insights into teachers' profiles. However, these questions require thorough revision to ensure the validity of the questions which adds to the workload and necessitates the involvement of domain experts.

Therefore, this approach would benefit from further refinement. One key area for improvement is that the data were single sourced in this study, which avoided issues related to that data fusion process. In real-word applications, these data often come from diverse platforms requiring careful integration. Additionally, some teacher related data must be collected through validated questionnaires to ensure accuracy and comprehensiveness and aspect that should be addressed in future improvements.

5 CONCLUSIONS

Data heterogeneity within gamified e-learning systems is quite challenging for ontology-based recommender systems. Teachers, learners, resources, and game elements typically originate from multiple sources, ranging from learning management systems and educational repositories to questionnaires and APIs, where each has varying formats and inconsistent semantics. Thus, driving our main research question for this study, that is how to collect and integrate heterogenous data from multiple sources for ontology−based models in gamified e−learning contexts. For that, an approach was proposed to manage the complexities of collecting and preparing heterogeneous data in ontology-based recommender systems for the gamified e-learning contexts. It proposes automated techniques to collect data from various sources, like Moodle and Coursera and maps them to our proposed Teacher in Gamified e-learning Context (TGC) ontology. Future research directions include development and refinement of semi-automated data collection methods, such as structured questionnaires, to gather comprehensive information about teachers' playing style, teaching style, and pedagogical and gamification experience. These questionnaires could be validated through pilot testing with representative teacher samples to ensure content validity and reliability. Data collection in this regard will be validated through a semi-automated process to ensure completeness and conciseness. Particular attention should be given to ensuring the content validity and reliability of these questionnaires, as well as the systematic validation of the collected data to enhance the accuracy and relevance of recommendations provided by ontology-based recommender systems. Thus, the development of such ontology-based recommender systems utilizing these integrated data is essential to provide personalized recommendations that align with teachers' needs and enhance the gamification process in e-learning environments. This involves leveraging the ontology to infer meaningful relationships between teacher profiles, resources, and game elements.

REFERENCES

Abel, E., Benayache, A., Lenne, D., & Moulin, C. (2007). *E-MEMORAe: A Content-Oriented Environment for E-Learning*. http://www.altrc.org/specification.asp.

Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., & Fernández-Manjón, B. (2020). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning*, *36*(3), 350–358. https://doi.org/10.1111/jcal.12405

Bakhouyi, A., Dehbi, R., Banane, M., & Talea, M. (2019). A semantic web solution for enhancing the interoperability of E-learning systems by using next generation of SCORM specifications. *International Journal of Emerging Technologies in Learning*, *14*(11), 174–185. https://doi.org/10.3991/IJET.V14I11.10342

Bennani, S., Maalel, A., Ghezala, H. Ben, & Daouahi, A. (2024). Towards an Adaptive Gamification Recommendation Approach for Interactive Learning Environments. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 199, pp. 341–352). Springer Science and Business Media Deutschland GmbH. https://doi.org/ 10.1007/978-3-031-57840-3_31

Bouihi, B., & Bahaj, M. (2019). Ontology and rule-based recommender system for e-learning applications. *International Journal of Emerging Technologies in Learning*, *14*(15), 4–13. https://doi.org/10.3991/ ijet.v14i15.10566

Del Blanco, A., Serrano, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2013). E-Learning standards and learning analytics. Can data collection be improved by using standard data models? *IEEE Global Engineering Education Conference, EDUCON*, 1255–1261. https://doi.org/10.1109/ EduCon.2013.6530268

Demertzi, V., & Demertzis, K. (2020). *A Hybrid Adaptive Educational eLearning Project based on Ontologies Matching and Recommendation System*.

Felder, R. M., & Spurlin, J. (2005). *Applications, Reliability and Validity of the Index of Learning Styles**.

Gomaa, Y., Moussa, S., Lahoud, C., & Abel, M. H. (2024). Exploring Recommender Systems for Assisting Teachers in E-Learning Gamification. *Procedia Computer Science*, *246*(C), 2312–2321. https://doi.org/ 10.1016/j.procs.2024.09.556

Joy, J., Raj, N. S., & Renumol, V. G. (2021). Ontology-based E-learning Content Recommender System for Addressing the Pure Cold-start Problem. *Journal of Data and Information Quality*, *13*(3). https://doi.org/ 10.1145/3429251

Jumaa, Y. M., Moussa, S. M., & Khalifa, M. E. (2017). *The Main Aspects of Adaptive Educational Games for Normal and Disabled / Disordered Learners : A Comprehensive Study*. *Icicis*, 348–355.

Maher, Y., Moussa, S. M., & Khalifa, M. E. (2020). Learners on Focus: Visualizing analytics through an integrated model for learning analytics in adaptive gamified e-learning. *IEEE Access*, *8*, 197597–197616. https://doi.org/10.1109/ ACCESS.2020.3034284

Mansouri, K., Kaiss, W., & Poirier, F. (2022). Personalized e-learning recommender system based on a hybrid approach. *IEEE Region*, *8*. https://doi.org/10.1109/ EDUCON52537.2022.9766650ï

Masud, M. (2016). Collaborative e-learning systems using semantic data interoperability. *Computers in Human Behavior*, *61*, 127–135. https://doi.org/10.1016/ j.chb.2016.02.094

Nashed, N. N., Lahoud, C., & Abel, M. H. (2021). TCO : a Teacher Context Ontology. *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021*, 757–762. https://doi.org/10.1109/CSCWD49262.2021. 9437622

Nashed, N. N., Lahoud, C., Abel, M.-H., & ene Abel, M. (2022). *Ontology-based teacher-context data* *integra-tion*. https://hal.archives-ouvertes.fr/hal-03537280

Nymfodora-Maria Raftopoulou, & Petros L. Pallis. (2023). Gamified Learning Systems' Personalized Feedback Report Dashboards via Custom Machine Learning Algorithms and Recommendation Systems. *Sociology Study*, *13*(3). https://doi.org/10.17265/2159-5526/2023.03.006

Otoo-Arthur, D., & van Zyl, L. , T. (2020). *A Scalable Heterogeneous Big Data Framework for e-Learning Systems*. IEEE.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The gamification user types Hexad scale. *CHI PLAY 2016 - Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 229–243. https://doi.org/10.1145/2967934.2968082

Villalba-Condori, K., Vera-Sancho, J., Vera-Vasquez, C., Mamani-Calcina, J., & Cardona-Reyes, H. (2022). *Knowledge-Based Recommender System in a Gamified Environment for Computational Thinking Development*.