International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-1794
Article

A Data-Driven Framework for Metaverse-Based Chinese Learning System Design Based on Knowledge Flow and Open Innovation

Xin Liu[1] and Pitipong Yodmongkol[1,*]

[1] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
* Correspondence author: liusubmission@163.com

Abstract: The emergence and fast growth of metaverse technologies have brought new possibilities to the field of immersive and interactive learning system architectures. Nonetheless, the current research is mainly concerned with the technological affordances and learning outcomes, and less emphasis is placed on the systematic organization of learning spaces in the metaverse. A solution to the gap is provided by this study which develops a data-driven framework of metaverse learning space design depending on knowledge flow and open innovation. It is a multi-stage methodology that combines practitioner knowledge, expert validation, and large scale data generated by users. In the first place, more than 11,000 records were gathered through Python-based web scraping of online platforms to find out patterns of learning activities. Secondly, five Chinese language teachers took part in structured discussions to interpret and refine these patterns into pedagogically meaningful categories. Lastly, expert validation was performed to confirm conceptual consistency. The knowledge management lifecycle, which was expanded to cover knowledge storage, was taken as an analytical framework. The findings indicate that knowledge acquisition (0.345) and practice (0.244) are prevalent in modern learning system architectures, whereas knowledge storage (0.170) is an important factor, and knowledge creation (0.060) is not well represented. These results led to a functional architecture of metaverse learning system architectures being created, which included the lecture, activity, materials storage, and assessment functions. Both theory and practice are enriched by this study because it (1) expands the knowledge management lifecycle, (2) introduces a quantitative modeling approach to the concept of knowledge flow, and (3) provides a systematic way of converting knowledge processes into learning space design. The suggested framework will provide a scalable solution to creating adaptive and knowledge-based metaverse learning system architectures.

Keywords: Metaverse learning systems; Knowledge flow modeling; Information system design; Open innovation; Data-driven framework

1. Introduction

Metaverse has come up as a new digital phenomenon combining the virtual reality (VR) and augmented reality (AR) technologies and networked social settings to form a continuous and immersive virtual world. It has gained more popularity in educational field over the past few years because it has a possibility to facilitate interactive, experiential and collaborative learning [1-2]. The metaverse-based environments allow learners to participate in embodied interaction and spatially situated activities, which are outside the traditional online learning system, and extend the frontiers of digital learning [3].

Although there were these improvements, the current research on metaverse-based education mostly concentrates on technical opportunities, user experience, and learning effectiveness. The previous researches discussed the architecture of immersive learning systems, virtual simulation, and outcomes of engagement but they offer little direction in regards to the structural organization of learning systems architecture within the metaverse [4-5]. Specifically, the problem of what functional learning spaces need to be built in the metaverse has not been adequately answered. This is a critical constraint since metaverse learning system architecture is not simply a virtual image of a real-world classroom but a modular system that consists of various functioning modules. Metaverse environments do not rely on predefined spatial structures, as in traditional learning system architectures, but require designers to actively decide the categories of learning spaces, i.e., instruction, interaction, practice, and assessment spaces, and their interrelationships. Nonetheless, the available literature mainly reflects on virtual classrooms and immersive experiences theoretically, but does not offer a systematic or evidence-based model of how these structures could be derived [2]. Information system viewpoint of this problem could be explained by the absence of mechanisms to convert learning processes into system structure. Literature on knowledge management offers a helpful background to this gap since it sees learning as a dynamic process involving the acquisition, sharing, creation and application of knowledge ([6-7]. These processes which are usually called as knowledge flow gives a systematic form of how learning activities are organized. Nevertheless, knowledge flow has not been sufficiently operationalized into spatial forms in metaverse settings, which leads to the separation of the theoretical models of learning and practical system design.

Moreover, the existing strategies of creating the learning system architecture also do not have a data-driven angle. Although big data and learning analytics have been extensively utilized to examine the behavior of learners and enhance educational results [8-9], its use has mainly concentrated on prediction, personalization and assessment. There has been little work utilizing large scale user generated data to shape the structural design of learning system architectures especially in new environments like the metaverse. This is a lost opportunity because online platforms are full of detailed descriptions of real-life learning experiences that can be applied to define trends in learning scenarios and activities. Moreover, the current literature is very scarce on integrating various sources of knowledge into a structured design process. Open innovation theory stresses that successful solutions may be achieved through the combination of external data, practitioner experience, and expert knowledge [10-11]. Nevertheless, today's metaverse research does not often integrate all three of these elements: big data, teacher input, and expert validation, in an integrated system architecture design of learning.

These drawbacks demonstrate a strong research gap: although the metaverse has become an extensively researched concept as a means of education, there is currently not a methodical, information-based, and knowledge-based way of deciding what types of learning spaces ought to be developed and how these spaces can be designed in metaverse settings. Three particular challenges have not been overcome:

(1) The lack of structured frameworks for identifying functional learning spaces in the metaverse;

(2) The lack of operationalization of knowledge flow into spatial configurations;

(3) The lack of integration of multi-source data and knowledge in the design process.

In order to overcome these issues, this paper suggests a data-driven architecture of design of metaverse learning environments which are dependent on knowledge flow and open innovation. In comparison to using predetermined categories of virtual spaces, the suggested framework develops learning space structures in several stages. Initially, massive amounts of user-generated data are gathered with the help of Python-based web scraping methods to discover trends of the current structure of learning systems. Secondly, Chinese language teachers engage in organized discussions to convert these trends into pedagogically valuable learning functions. Thirdly, expert validation is performed to improve and validate the final design. Knowledge flow is therefore systematically mapped to learning functions and then to metaverse learning spaces by this means.

Metaverse learning environments are defined in this study as an information system with functional modules that are motivated by knowledge flow. In accordance with that, the present study attempts to answer the following research question:

In what way can metaverse learning environments be implemented in a systematic manner and what kind of functional rooms should be built depending on the flow of knowledge?

The contributions of this study are threefold.

First, it provides a data-driven approach to identifying and structuring metaverse learning spaces.

Second, this creates a theoretical relation between knowledge management procedures and design of virtual space, which converts the flow of knowledge into spatial forms.

Lastly, it suggests an open innovation-based design framework that will allow integrating the big data, knowledge of practitioners and expert validation into a single and repeatable system.

2. Literature Review

The accelerated growth of metaverse technologies has made possible the creation of new architectures of immersive and interactive learning systems. The metaverse has been more and more understood as a continuous virtual environment which allows the embodiment of interaction, real-time interaction, and experiential learning by incorporating virtual reality (VR), augmented reality (AR), and networked infrastructures [1-2]. In an educational setting, these features have been linked to increased engagement, presence, and to the ability to simulate complex learning situations [3], [5]. Consequently, the metaverse has frequently been seen as a new generation learning system architecture, which goes beyond both conventional online and hybrid systems.

Nevertheless, even though metaverse-based education is attracting more and more interest, the current literature is mostly devoted to technological affordances and learning outcomes, not to the systematic design of learning system architectures. Many studies are based on the concept of virtual classrooms, immersive simulations or social interaction space [4], [12], with few directions on how the internal structure of learning system architectures ought to be organized. Specifically, there are no methodological models of how functional learning spaces should be built within the metaverse, and how these spaces should be arranged to facilitate various learning processes. To resolve this problem, it is essential to take into account the function of knowledge processes in the design of learning system architecture. The theory of knowledge management offers a systematic view of how knowledge is created, shared, and used. The knowledge management lifecycle conceptualizes knowledge as a dynamic process, which comprises the steps of acquisition, sharing, creation, application, and validation [6-7], [13]. These stages in the education setting can be seen as a reflection of different forms of learning activities and interactions, and it can be suggested that learning system architectures can be developed as systems that facilitate certain knowledge processes. This gap suggests that the challenge of designing metaverse learning system architectures is not only technological, but also pedagogical and process oriented. Although metaverse environments offer immersive and interactive affordances, the educational value of such environments depends on how learning activities are organized and how different forms of knowledge engagement are supported within the system [3], [14]. From a learning design perspective, a meaningful learning system architecture cannot be defined solely by virtual presence or interface features; rather, it must be structured around the processes through which learners acquire, share, apply, and construct knowledge [15-16]. This makes knowledge processes a necessary analytical bridge between technological environments and functional learning space design.

Although it is relevant, the knowledge management lifecycle has not been implemented operationally in the development of metaverse learning system architecture. The available literature mostly relies on knowledge management as an explanatory model, but not as a designer tool to organize learning system architectures. Specifically, there is a dearth of studies on how information flow may be converted into spatial arrangements, like identification of various kinds of virtual spaces depending on the information processes. This gap indicates that the conceptual understanding should be shifted to the system level implementation. Simultaneously, learning system architecture design is now becoming more dependent on the combination of a wide variety of knowledge sources. Open innovation theory suggests that successful solutions come out with combination of outside data, practitioner experience and expert knowledge [10-11]. When designing a complex system, including a metaverse context, it is not enough to have access to only one source of knowledge. To address this, a systematic approach is required to synthesize data-related knowledge, experiential knowledge and theoretical knowledge.

However, existing research on metaverse learning system architectures rarely adopts an open innovation perspective. Most studies either rely on conceptual design approaches or focus on specific user groups, without integrating big data, practitioner knowledge, and expert validation into a unified design framework. This limitation reduces the robustness and generalizability of proposed learning system architecture models. In addition, although data-driven approaches have become increasingly important in education, their application remains limited in the context of learning system architecture design. Learning analytics research has demonstrated the potential of large-scale data to reveal patterns in learner behavior and interaction [8-9], yet these approaches are primarily used for prediction and personalization rather than for structural design. The use of user-generated data to inform the configuration of learning system architectures, particularly in the metaverse, remains underexplored.

Knowledge management theory provides a structured perspective for understanding how knowledge is generated, shared, and applied. The knowledge management lifecycle conceptualizes knowledge as a dynamic process involving stages such as acquisition, sharing, creation, application, and validation [6-7], [13]. In educational contexts, these stages correspond to different types of learning activities and interactions, suggesting that learning system architectures can be designed as systems that support specific knowledge processes. This theoretical relevance suggests that the knowledge management lifecycle has the potential to serve not only as an interpretive lens, but also as a design logic for structuring learning system architectures. If different learning activities reflect different knowledge processes, then functional learning spaces may also be designed to correspond to those processes. In principle, this provides a pathway for translating abstract knowledge flow into concrete system structures. However, this potential has not yet been fully realized in existing metaverse learning research.

Nevertheless, the current body of literature pertaining to the metaverse learning system architectures hardly incorporates open innovation view. Many of the works are based on the conceptual design approach or they deal with particular user groups, but do not incorporate big data, practitioner knowledge, and expert validation as part of the overall design framework. This constraint limits the strength and applicability of the suggested learning system architecture designs. Moreover, in spite of the fact that data-driven methods have gained prominence in the field of education, they are still not widely applied to the sphere of learning system architecture design. The study of learning analytics has shown that massive amounts of data could be used to uncover trends in the behavior and interactions of learners [8-9] but these methods are mainly utilized in the prediction and customization processes instead of the structural design process. There is still much uncharted territory in the area of using user-created information to influence the design of learning system architectures, especially in the metaverse.

The literature shows that there is a gap in the intersection of three fields, which are the metaverse learning system architectures, knowledge management and open innovation. Although the metaverse technologies bring fresh opportunities to learn, knowledge management has a theoretical base to comprehend the learning procedures, and open innovation has a solution to integrate various sources of knowledge, these views have never been integrated in a systematic way. Hence, there is a necessity to create a framework that (1) applies the knowledge management lifecycle as a starting point of the learning system architecture design, (2) uses data-driven approaches to determine learning patterns, and (3) combines various sources of knowledge using an open innovation process. The current research satisfies this gap by introducing a data-driven framework that turns knowledge flow into functional metaverse learning spaces offering a systematic and replicatable method to learning system architecture design.

3. Methodology

The theoretical framework that is used in this study is theory-driven and data-driven design methodology in developing a framework of metaverse learning space design. The paper is based on two fundamental methodological principles - the knowledge management lifecycle and the open innovation process. The knowledge management cycle is the analytical structure used in understanding learning activities as knowledge-based processes, whereas the open innovation process gives the steps to follow when integrating various sources of knowledge. Instead of presupposing a set learning space, it will be used to determine how to approach the translation of empiric data and knowledge processes into design dimensions, which can be used in metaverse learning system architectures. Therefore, this research is designed as a design-based framework development research where large-scale data analysis and practitioner interpretation are integrated to form a structured approach to learning space design.

Knowledge management lifecycle can be considered as the main analytical model in this work. Knowledge acquisition, dissemination, identification, refining, creation, and validation are knowledge-related processes that are used to interpret learning activities. The processes offer a systematic foundation on which to arrange patterns derived by the data and comprehend the functional roles of learning activities. Significantly, the lifecycle is applied as a methodological instrument not as an abstract reference. It directs (1) formulating analytical questions, (2) interpreting extracted keywords, and (3) arranging learning activities into functional dimensions. By this process, data patterns that are descriptive become theoretical representations of learning processes. In the present research, the knowledge management lifecycle is broadened to involve knowledge storage as a separate step. The lifecycle thus comprises six interconnected elements, i.e., knowledge acquisition, knowledge sharing, knowledge storage, knowledge practice, knowledge creation, and knowledge validation. Knowledge storage is especially significant in digital and metaverse-based learning system architecture that is characterized by the presence of persistent learning resources, reusability and access to such resources in various learning environments. In contrast with the traditional classroom setting, the metaverse environment needs to have an organized system of storage and retrieval of learning materials, so knowledge storage becomes one of the essential elements of the learning process.

The present study also uses an open innovation approach as the process mechanism to integrate knowledge. Open innovation in this work is operationalized as a systematic process that incorporates three knowledge sources, which are (1) external knowledge based on large scale online data, (2) practitioner knowledge that is made up by teachers, and (3) theoretical direction offered by the knowledge management lifecycle.

The research procedure was organized into three sequential stages. In the first stage, large-scale user-generated data were collected from online platforms through Python-based web scraping. These data were then cleaned and analyzed to identify recurrent patterns related to learning activities, interaction modes, and learning contexts. In the second stage, practitioner knowledge was incorporated through structured discussion with Chinese language teachers, who reviewed the extracted keywords, interpreted their pedagogical relevance, and supported the categorization of learning activities. In the third stage, the categorized activity patterns were mapped onto the knowledge management lifecycle and transformed into corresponding learning functions and metaverse learning spaces. Expert validation was then conducted to examine the conceptual consistency and practical relevance of the proposed structure.

To establish a data-driven foundation, large-scale user-generated data were collected from online platforms using Python-based web scraping. The primary data source was Zhihu, a Chinese question-and-answer platform containing extensive discussions related to learning experiences and educational practices. Although the dataset may introduce cultural bias due to platform-specific characteristics, the inclusion of multiple sources—Zhihu, WeChat, and Facebook—partially mitigates this limitation. In total, 7,346 Zhihu records were collected, of which 6,916 were retained after data cleaning, resulting in a usability rate of approximately 94%. In addition, 2,701 records from WeChat and 2,201 records from Facebook were collected as supplementary practitioner-oriented data, yielding a total of 4,902 records. Together, these datasets provide complementary perspectives by combining general user experiences with teacher-related discourse.

A large-scale user-generated data were gathered through web scraping of online platforms with Python to create a data-driven base. The main data source was Zhihu, a Chinese question-and-answer site with numerous discourses pertaining to learning experiences and educational practices. The given dataset might have some cultural bias because of the specific features of the platforms. Nevertheless, the fact that several sources are used (Zhihu, WeChat, Facebook) somewhat alleviates this weakness. The future studies could use the framework with cross-cultural data. This research has collected 7,346 records in total and 6,916 records remained after the data cleaning process which means the rate of usability is around 94%. As a supplement to this dataset, other practitioner-centered data were gathered using social media platforms, namely, 2,701 records on WeChat and 2,201 records on Facebook, which results in the overall number of 4,902 records. Both of these datasets offer different points of view by integrating overall user experiences with discourse about teachers. Besides external data sources, the study included five Chinese language teachers in Chiang Mai as practitioner participants. They were chosen according to their participation in the work of teaching Chinese language and their ability to interpret the learning activity in a real-life context of education. Their purpose was not to present statistically meaningful information, but to facilitate contextual understanding and development of knowledge. The participating teachers discussed the extracted keywords by engaging in interpreting them, assessing their relevance to the teaching practice, and helping in the classification of learning activities. Herein, they served as internal domain experts in the open innovation process, as they bridged the gap between large scale data and real world educational practice.

The overall system architecture of the metaverse learning space design is illustrated in Figure 1 based on this transformation pipeline. The illustration shows how raw data are converted into metaverse learning spaces by using feature extraction, modeling of knowledge flows, and functional mapping.

Phase 1: Data Acquisition and Feature Extraction

In the first phase, large-scale user-generated data are collected from online platforms using Python-based web scraping techniques. The dataset is defined as:

$$D = \{d1,\ d2,\ ...,\ dn\} \tag{1}$$

These data consist of textual descriptions of learning experiences, learning situations, and activity patterns. A feature extraction function is applied to identify relevant patterns:

$$F = g(D) \tag{2}$$

D represents the raw textual data collected from online platforms, including user discussions on Zhihu, WeChat, and Facebook related to learning experiences and educational practices.

$g(D)$ represents the process applied to the raw data. In this study, this process refers to term frequency–inverse document frequency–based feature extraction, which is used to identify representative high-frequency words and meaningful keywords from the textual data.

F represents the extracted features obtained after this processing step. These features consist of learning-related keywords, such as classroom, study, knowledge, activity, communication, and test, which reflect recurring patterns of learning activities, interaction modes, and learning contexts within the dataset.

Phase 2: Knowledge Flow Modeling

The extracted features are mapped to structured knowledge components based on the knowledge management lifecycle:

$$K = h(F) \tag{3}$$

$h(F)$: mapping via rule-based classification

F represents the set of extracted features obtained from the textual data, including learning-related keywords such as classroom, study, knowledge, activity, communication, and test.

$h(F)$ represents the knowledge mapping process applied to these features. In this study, this process refers to classifying the extracted keywords into corresponding knowledge processes based on the knowledge management lifecycle.

K represents the resulting knowledge-process categories. These categories include knowledge acquisition, knowledge sharing, knowledge storage, knowledge practice, knowledge creation, and knowledge validation.

The knowledge flow is defined as:

$$K = \{K_a, K_s, K_{st}, K_c, K_{ap}, K_v\} \tag{4}$$

where:

- K_a : *acquisition*
- K_s : *sharing*
- K_{st} : *knowledge storage*
- K_c : *creation*
- K_{ap} : *application/practice*
- K_v : *validation*

To quantify the relative importance of each knowledge component, frequency-based intensity is calculated as:

$$I(K_i) = f_i / \sum f_j \tag{5}$$

where f_i denotes the frequency of occurrence of knowledge component K_i.

$I(K_i)$ represents the normalized intensity of knowledge process i.

f_i denotes the frequency of occurrence of features associated with knowledge process i, and $\sum f_j$ represents the total frequency of all knowledge processes across the dataset.

This formulation measures the proportion of each knowledge process relative to the overall distribution, allowing comparison of their relative prominence in the data.

Phase 3: Open Innovation–Based Knowledge Integration

To enhance the robustness of the design process, the study integrates multiple knowledge sources through an open innovation approach. Three sources are considered:

- Data-driven insights (*D*)
- Teacher input (*T*)
- Expert validation (*E*)

The normalized weights for teacher and expert inputs are defined as:

$$w\_T(K_i) = t_i / \sum t_j \tag{6}$$

$$w\_E(K_i) = e_i / \sum e_j \tag{7}$$

$w\_T(K_i)$ represents the weight of knowledge process i derived from teacher input.

t_i denotes the frequency of teacher-related features associated with knowledge process i, and $\sum t_j$ represents the total frequency of all teacher-related features across knowledge processes.

Similarly, $w\_E(K_i)$ represents the weight of knowledge process i derived from expert input.

e_i denotes the frequency or assigned value associated with knowledge process i based on expert validation, and $\sum e_j$ represents the total across all knowledge processes.

Both formulations use normalization to ensure that the weights are comparable across different knowledge sources.

The overall score of each knowledge component is calculated as:

$$Score(K_i) = \alpha I(K_i) + \beta w\_T(K_i) + \gamma w\_E(K_i) \tag{8}$$

$Score(K_i)$ represents the final score of knowledge process i.

$I(K_i)$ denotes the normalized intensity derived from large-scale data analysis.

$w\_T(K_i)$ represents the normalized weight based on teacher input.

$w\_E(K_i)$ represents the normalized weight based on expert validation.

The weighting process follows a normalized multi-criteria decision model, this formulation is conceptually aligned with multi-criteria decision-making (MCDM) approaches such as AHP. The proposed model enables reproducible computation of knowledge flow distribution from raw data inputs.

subject to:

$$\alpha + \beta + \gamma = 1 \tag{9}$$

This formulation enables the integration of quantitative data, practitioner knowledge, and expert judgment into a unified framework.

Phase 4: Learning Function and Space Mapping

Each knowledge component is transformed into a corresponding learning function:

$$LF_i = f(K_i) \tag{10}$$

where K_i represents a specific knowledge process (e.g., acquisition, practice, or validation), and LF_i denotes the corresponding learning function derived from that process (e.g., lecture, activity, or assessment). Here, f represents a conceptual mapping based on pedagogical interpretation rather than a strictly defined mathematical function.

Subsequently, learning functions are mapped to metaverse learning spaces:

$$R_i = \varphi(LF_i) \tag{11}$$

where R_i represents a metaverse learning space designed to support learning function LF_i, and $\varphi(\cdot)$ denotes the transformation from functional design to spatial implementation within the metaverse environment.

This establishes a direct mapping from knowledge flow to spatial design:

$$R_i = \varphi(f(K_i)) \tag{12}$$

This formulation indicates that metaverse learning spaces are derived from knowledge processes through an intermediate functional layer, rather than being predefined independently of learning logic.

The complete set of learning spaces is defined as:

$$R = \{R_1, R_2, \ldots, R_n\} \tag{13}$$

where each space corresponds to a specific function grounded in the knowledge management lifecycle.

The overall research process consists of three operational steps:

(1) Data Mining: Python-based web scraping is used to collect and analyze user-generated learning data.

(2) Teacher Co-Design: Chinese language teachers interpret and refine data-derived patterns into pedagogically meaningful learning functions.

(3) Expert Validation: Experts evaluate and validate the proposed learning space structure to ensure conceptual consistency and practical relevance.

This process reflects an open innovation mechanism in which knowledge is progressively refined through multiple layers of input.

Based on this transformation pipeline, the overall system architecture of the metaverse learning space design is illustrated in Figure 1. This figure illustrates how raw data are transformed into metaverse learning spaces through feature extraction, knowledge flow modeling, and functional mapping.

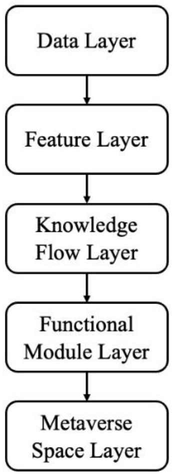

Figure 1. Information System Architecture of Metaverse Learning Space Design

3. Research Results

3.1. Feature Extraction from Big Data

To identify empirical patterns of learning system architectures, a large-scale dataset was analyzed, including:

- 6,916 valid Zhihu records
- 4,902 teacher-related records (WeChat + Facebook)

Frequency analysis was conducted to extract key features:

$$F = g(D) \tag{14}$$

The most frequent keywords include:

Acquisition

- classroom (5912)
- study (5847)
- knowledge (5631)
- lecture (5242)
- theory learning (1880)

Sharing

- communication (3192)
- question answering (2558)
- group learning (1894)

Storage

- learning materials (2359)
- video (2056)
- reading (1975)

Creation

- research (1867)

Practice

- activity (3849)
- daily life (3352)
- technology (3116)
- home (3012)
- self-learning (2922)

- language culture (2880)

Validation

- test (2420)

These keywords represent recurring patterns in learning activities and environments.

3.2. Knowledge Flow Modeling

The extracted features were mapped to knowledge processes based on the knowledge management lifecycle:

$$K = h(F) \tag{15}$$

$$K = \{K_a,\ K_s,\ K_{st},\ K_c,\ K_{ap},\ K_v\} \tag{16}$$

where:

- K_a *: acquisition*
- K_s *: sharing*
- K_{st} *: knowledge storage*
- K_c *: creation*
- K_{ap} *: application/practice*
- K_v s*: validation*

3.3. Quantification of Knowledge Flow

To quantify the relative importance of each knowledge process, frequency-based intensity was calculated:

$$I(K_i) = f_i / \sum f_j \tag{17}$$

Based on keyword distribution and mapping, the normalized intensity values are (Table 1):

- **Acquisition**: classroom, study, knowledge, lecture, reading, video, learning materials, theory learning
- **Sharing**: communication, question answering, group learning
- **Practice**: activity, daily life, technology, home, self-learning, language culture
- **Creation**: research
- **Validation**: test

Table 1. Normalized Intensity of Knowledge Processes Based on User-Generated Data

Knowledge Process	$I(K_i)$
Acquisition	0.389
Sharing	0.121
Storage	0.101
Practice	0.304
Creation	0.030
Validation	0.038

The results indicate that **acqisition-related processes have the highest intensity**, while creation processes have relatively lower representation.

3.4. Open Innovation–Based Weight Integration

To improve robustness, practitioner and expert inputs were incorporated.

Teacher-based weight (Table 2):

Derived from 32 teacher interviews and 464 CoP records

$$w\_T(K_i) \tag{18}$$

- **Acquisition**: classroom, study, knowledge, lecture
- **Sharing**: community, question answering
- **Storage:** resource(book), video, learning materials

- **Practice**: activity, technology, language culture
- **Creation**: no direct keyword in current teacher table
- **Validation**: test

Table 2. Teacher-Based Weight Distribution of Knowledge Processes

Knowledge Process	$w\_T(K_i)$
Acquisition	0.432
Sharing	0.144
Storage	0.266
Practice	0.183
Creation	0.000
Validation	0.075

Expert-based weight (Table 3):
Derived from 4 expert validation participants

$$w\_E(K_i) \tag{19}$$

Table 3. Expert-Based Weight Distribution of Knowledge Processes

Knowledge Process	$w\_E(K_i)$
Acquisition	0.167
Sharing	0.167
Storage	0.167
Practice	0.167
Creation	0.166
Validation	0.166

The final score of each knowledge process is calculated as (Table 4):

$$Score(K_i) = \alpha I(K_i) + \beta w\_T(K_i) + \gamma w\_E(K_i) \tag{20}$$

$$\alpha + \beta + \gamma = 1 \tag{21}$$

Table 4. Integrated Scores of Knowledge Processes Based on Multi-Source Weighting

Knowledge Process	Final Score
Acquisition	0.345
Sharing	0.143
Storage	0.170
Practice	0.244
Creation	0.060
Validation	0.089

The inclusion of knowledge storage as a separate activity makes the allocation of knowledge elements more equitable. The acquisition of knowledge is still the most prevalent process (0.345), then comes practice (0.244) and knowledge storage (0.170) which are important factors in the instructional input and continuous learning resources. The addition of knowledge storage to the previous five-component model decreases the hegemony of acquisition, and it emphasizes the significance of the availability of resources in the architecture of metaverse learning systems. This change offers a better depiction of the learning process in a digital and virtual environment.

However, these findings are not of an optimal structure, but they show the structural shortcomings of the existing learning system designs, which is a very strong reason to reconstruct the learning spaces to be more even in knowledge processes.

The functional organization of the learning system architecture had not been pre-determined but it was obtained through the weighted distribution of the knowledge processes. In particular, those processes that scored the highest were converted into leading functional parts whereas those with a lower score were depicted as auxiliary or supplementary sub-functional parts.

The weighted distribution of knowledge processes is used to derive the functional architecture. Acquisition is the main instructional layer whereas practice is the underlying activity layer. Creation, a low-weight process, is depicted as an extended learning element. The functions that support it, such as materials storage and assessment, are performed in all layers to enable the dissemination of knowledge and its verification.

Table 5 shows the mapping between knowledge processes and learning system architecture functions when knowledge storage is added as a separate element. The findings indicate that knowledge storage also turns out to be a crucial process (0.170) besides acquisition and practice, which underscores the value of permanent and available educational materials in the metaverse settings.

Table 5. Mapping from Knowledge Processes to Learning system architecture Functions

Knowledge Process	Score	Mapped Function	Interpretation
Acquisition	0.345	Lecture Function	structured knowledge input
Practice	0.244	Activity Function	active application and skill use
Storage	0.170	Materials Storage Function	persistent and accessible learning resources
Sharing	0.143	Communication / Group Work	interaction and collaborative exchange
Validation	0.089	Assessment Function	learning confirmation and feedback
Creation	0.060	Self-learning / Culture-related extension	higher-order and contextual knowledge development

With this mapping, it can be assured that every functional element of the learning system architecture is established based on an equivalent knowledge process and hence a systematic conversion between the flow of knowledge and functional design is achieved.

The weighted distribution of knowledge processes was later converted into a functional learning-environment structure. The highest-ranked process, Acquisition, was turned into the Lecture Function, which means that structured knowledge input is still relevant. The next highest process, Practice, served as the heart of the Activity Function and Sharing was operationalized in the form of communication and collaboration oriented sub-functions. Validation was linked to the Assessment Function, and Creation was expressed through self-study and cultural extensions. Using these assumptions, the quantified process distribution was transformed into the functional architecture of the learning system architecture.

According to the quantified distribution of knowledge and the suggested target structure, learning system architecture functions have been recognized and interpreted in order to operationalize the framework. The given functions are a representation of the translation of knowledge processes into pedagogically significant elements of the architecture of the learning system.

Namely, knowledge acquisition refers to organized instructional functions whereas practice and sharing process refers to activity-oriented functions. Assessment functions are related to validation processes and supply of materials storage functions deals with resource-related needs. The final functional structure is given in Table 6.

Table 6. Functional Definitions of Learning system architecture Components Derived from Knowledge Flow

Learning system architecture Function	**Definition (in this study)**	**How the function is used in the learning system architecture**	**Related student learning needs addressed**
Lecture Function	A structured instructional function focused on systematic knowledge input and concept explanation, primarily led by the teacher.	Used at the beginning of learning cycles to introduce new language knowledge (e.g., vocabulary, grammar, pronunciation rules) and provide conceptual frameworks that support subsequent activities.	Supports students' need for clear learning guidance and structured content, which was rated at a high level in overall environment evaluation.
Activity Function	A learner-centered function that supports interactive, practice-oriented, and experiential learning through multiple sub-functions.	Implemented through group work, communication tasks, skill practice, self-learning tasks, and culture-based activities. Serves as the core space for applying, practicing, and extending language knowledge.	Directly responds to students' median-level needs for interaction, active participation, and sustained engagement, identified as insufficiently supported in the existing environment.
Group Work (sub-function)	A collaborative learning sub-function that enables learners to co-construct knowledge through peer interaction.	Used for pair or group-based language tasks such as dialogues, role-plays, problem-solving, and project work.	Addresses students' need for peer collaboration, which showed relatively lower perceived support in the existing learning system architecture.
Communication (sub-function)	A sub-function focused on real-time or asynchronous interaction between teachers and students and among peers.	Implemented through discussion spaces, feedback channels, and question–answer interactions during and after class.	Responds to students' need for teacher–student interaction and timely communication, which was rated at a median level.
Skill Practice (sub-function)	A practice-oriented sub-function targeting specific language skills.	Used for listening, speaking, reading, and writing practice activities, including drills, simulations, and task-based exercises.	Supports students' need for skill-based learning opportunities and active language use, contributing to sustained engagement.
Self-learning (sub-function)	A sub-function that enables learners to control learning pace, time, and content independently.	Implemented through self-paced tasks, supplementary learning modules, and independent practice activities.	Directly addresses students' median-level need for autonomous learning support, identified as insufficiently supported in the existing environment.
Culture Delivery (sub-function)	A sub-function integrating language learning with cultural and real-life contexts.	Used through culture-related tasks, daily-life scenarios, and contextualized language activities.	Responds to students' need for contextualized learning and meaningful engagement beyond formal classroom instruction.
Materials Storage Function	A resource-oriented function that provides centralized access to learning materials and learning support tools.	Used as a shared repository for learning materials, videos, readings, tasks, and references, accessible across all learning spaces.	Supports students' need for flexible access to learning resources, enabling learning beyond fixed class time and space.
Assessment Function	A validation-oriented function that evaluates learning outcomes and learning progress.	Used through formative and summative assessments aligned with learning activities, including quizzes, performance tasks, and reflective assessments.	Addresses students' need for clear learning outcomes and feedback, supporting learning effectiveness and self-monitoring.

5. Discussion and Conclusion

This proposed framework can be viewed as a modular information system design of metaverse based learning platforms. In this study, it is argued that there should be a data-driven framework to design metaverse learning system architectures through the integration of the knowledge management life cycle with an open innovation process. In contrast to conventional studies that perceive learning system architectures as fixed educational environments, this paper views learning system architectures as dynamic systems organized by knowledge flows, which is in line with the current views on digital learning ecosystems [14], [17]. The findings show that knowledge processes are not evenly spread throughout the structure, and knowledge acquisition (0.345) and practice (0.244) dominate the structure, whereas knowledge creation (0.060) is underrepresented. It is in agreement with other literature indicating that most technology-enhanced learning system architectures still copy traditional lecture-based pedagogies instead of facilitating higher-order learning [18-19]. Theoretically speaking, this paper extends the knowledge management lifecycle by adding knowledge storage as one of the independent processes. Traditional models of knowledge management usually focus more on acquisition, sharing, and creation [6-7] and pay less heed to storage as one of their main components. Nevertheless, in digital and metaverse-based environments, constant and accessible learning materials are essential to facilitate continuous learning across contexts [3], [14]. The addition of knowledge storage then gives a better overall picture of the knowledge flow within immersive learning system architectures. In the context of design, the results indicate how knowledge processes can be transformed into functional elements of learning system architectures in an organized way. This is in accordance with constructivist and activity based learning theories which suggest that interaction, collaboration and contextualized learning are all important [15]. The Activity Function, driven by knowledge practice, becomes the core element surrounded by sub-functions that are oriented towards sharing, including communication and group work. Meanwhile, the low representation of knowledge creation indicates a discrepancy between the existing practices of learning and the expectations of the metaverse, which are frequently linked to creativity, co-creation, and immersive involvement [1], [3].

It is important to note that this paper also uses an open innovation process to incorporate various sources of knowledge such as user generated data, inputs given by practitioners, and validation provided by experts. It is in line with the idea of open innovation theory, which stresses on the combination of internal and external knowledge to be able to solve problems more effectively [10]. Through the integration of mass data analysis and practitioner interpretation, the research contributes to the empirical basis and practical applicability of the proposed framework. The results indicate that the learning architecture of metaverse learning systems must not be merely based on traditional teaching structures. They ought to be developed as layered systems that have various knowledge processes spread over dedicated learning environments that allow a more flexible, interactive and student-oriented experience. The credibility of the framework is also reinforced by the congruence between the outcomes of data-driven analysis and practitioner insights. Even though the complete implementation of the system is not possible in this present study, the framework will be used to support future empirical verification.

The given structure has a high potential to be used in practice in various fields. It can be applied in intelligent tutoring systems in order to provide adaptive learning paths through knowledge flow modeling thus offering more individualized and data driven instructional design. The framework could be used in corporate training platforms to help structure training environments as modular learning spaces that relate to the various stages of knowledge acquisition, practice, and validation which consequently enhances training effectiveness and knowledge retention. Additionally, in the context of industrial simulation based on the metaverse, the framework offers a methodological way of creating immersive training settings in which the processes of gaining knowledge are integrated into interactive situations, facilitating experiential learning and skills acquisition of multifaceted real-life activities. These uses make the framework go beyond educational settings and make contributions to the creation of scalable and knowledge-oriented information systems in the academic environment and the industry.

The research creates a knowledge flow-based model of metaverse learning space design which fills the gap in the absence of systematic methods of designing learning system architectures in digital and immersive environments. The study has come up with a systematic process of data-driven analysis, theoretical modeling, and practitioner knowledge to create a structure that transforms data into knowledge processes, and then into functional design and ultimately into learning system architecture architecture.

The contributions of this study are threefold.

Then, it adds to the knowledge management lifecycle by including knowledge storage as a core process, which gives a better picture of learning within digital environment settings.

This is followed by its introduction of a quantitative way of modeling knowledge flow allowing the identification of dominant and underrepresented learning processes in terms of empirical data.

Finally, it suggests the functional architecture of metaverse learning system architectures which shows how knowledge processes may be operationalized in the learning space design, see Figure 2.

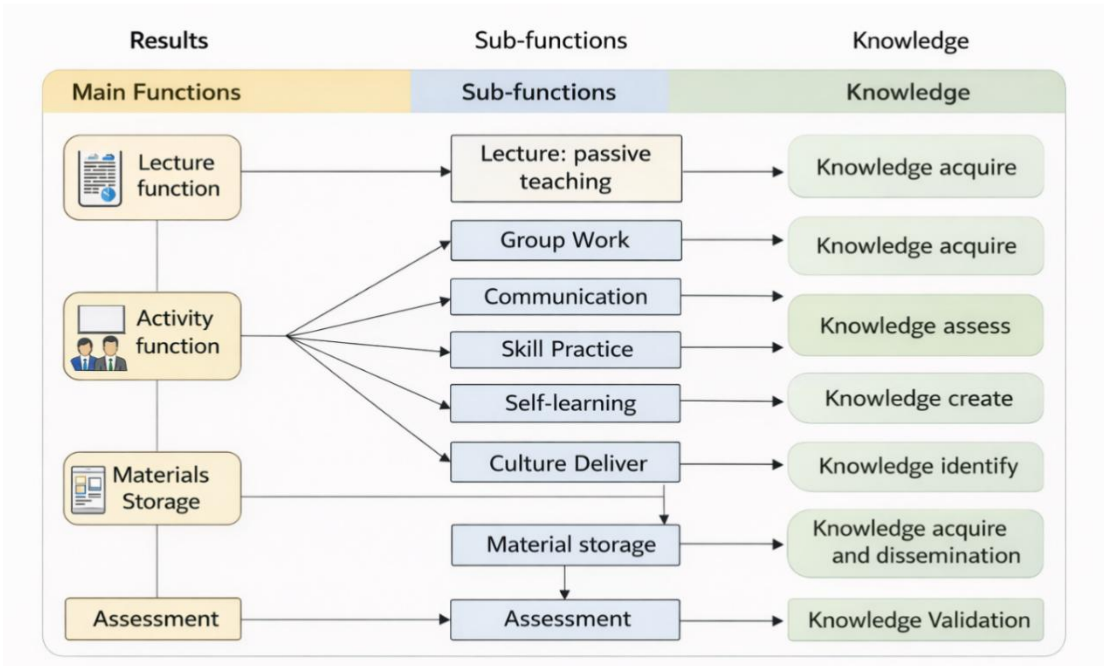

Figure 2. Metaverse learning system architecture diagram

Nevertheless, there are a number of constraints that can be considered. The research is based on information obtained through certain online platforms and a small group of practitioner respondents, which might have an impact on the generalizability. Moreover, the expert validation procedure was not completely measured in terms of individual knowledge processes, implying possibilities of more stringent rating methods in future studies.

The future direction of research could be to expand on the current study by adding a cross-cultural dataset, performing empirical validation in actual metaverse settings, and using more sophisticated modeling methods to continue investigating the connection between the flow of knowledge and the learning process.

To sum up, the paper switches the emphasis of metaverse learning system architecture design away from a technology-focused approach to a knowledge-based approach and gives them an organized and scalable framework to create more adaptive and effective learning system architecture designs in the age of immersive education.

Acknowledgement

The research was supported by CMU Presidential Scholarship.

Reference

1. Mystakidis, S. (2022). Metaverse. Encyclopedia, 2(1), 486–497.
2. Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., et al. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 66, 102542.
3. Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education. Computers & Education, 147, 103778.
4. Kye, B., Han, N., Kim, E., Park, Y., & Jo, S. (2021). Educational applications of metaverse: Possibilities and limitations. Journal of Educational Technology & Society, 24(3), 1–12.

5. Makransky, G., & Petersen, G. B. (2019). Immersive virtual reality and learning: A meta-analysis. Educational Psychology Review, 31, 1–22.
6. Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company. Oxford University Press.
7. Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems. MIS Quarterly, 25(1), 107–136.
8. Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK).
9. Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5–6), 304–317.
10. Chesbrough, H. (2003). Open innovation: The new imperative for creating and profiting from technology. Harvard Business School Press.
11. West, J., & Bogers, M. (2014). Leveraging external sources of innovation: A review of research on open innovation. Journal of Product Innovation Management, 31(4), 814–831.
12. Lee, L.-H., Braud, T., Zhou, P., et al. (2021). All one needs to know about metaverse: A complete survey on technological singularity. IEEE Access, 9, 153004–153020.
13. Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba, and leadership: A unified model of dynamic knowledge creation. Long Range Planning, 33(1), 5–34.
14. Dede, C. (2009). Immersive interfaces for engagement and learning. Science, 323(5910), 66–69.
15. Jonassen, D. H. (1999). Designing constructivist learning system architectures. In C. M. Reigeluth (Ed.), Instructional design theories and models (Vol. 2).
16. Carvalho, L., & Goodyear, P. (2018). Design, learning networks and service innovation. Springer.
17. Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1).
18. Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. Teachers College Press.
19. Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. Teaching and Teacher Education, 67, 135–142.