Virtual Patient Platform and Data Space for Sharing, Learning, Discussing, and Researching

Santanchè, Andre; Mattosinho, Heitor; de Menezes Mota, Marcos F.; Pantoja, Fagner Leal; de Freitas Leite, Gabriel; Tonelli, Ana; Salvetti Valente, Fernando; de Castro Solano Martins, Juliana; Queiros, Sandro; Grangeia, Tiago De Araujo Guerra

Published in:
IEEE International Conference on e-Science and Grid Computing

DOI:
10.1109/e-Science58273.2023.10254792

Virtual Patient Platform and Data Space for Sharing, Learning, Discussing, and Researching

André Santanchè
Institute of Computing
University of Campinas
Campinas, Brazil
santanche@ic.unicamp.br

Heitor Soares Mattosinho
Institute of Computing
University of Campinas
Campinas, Brazil
heitor.mattosinho@ic.unicamp.br

Marcos Felipe de Menezes Mota
Institute of Computing
University of Campinas
Campinas, Brazil
marcos.mota@ic.unicamp.br

Fagner Leal Pantoja
Institute of Computing
University of Campinas
Campinas, Brazil
fagner.pantoja@ic.unicamp.br

Gabriel de Freitas Leite
Faculty of Mechanical Engineering
University of Campinas
Campinas, Brazil
gabriel.dfleite@gmail.com

Ana Claudia Tonelli
Internal Medicine Service
Hospital de Clinicas de Porto Alegre (HCPA)
Porto Alegre, Brazil
actoliveira@hcpa.edu.br

Fernando Salvetti Valente
Department of Internal Medicine
HCFMUSP
São Paulo, Brazil
f.valente@hc.fm.usp.br

Juliana de Castro Solano Martins
Bahia School of Medicine
UFBa - Federal University of Bahia
Bahia, Brazil
julianacsmartins@gmail.com

Sandro Queirós
Life and Health Sciences Research Institute
University of Minho
Braga, Portugal
sandroqueiros@med.uminho.pt

Tiago de Araujo Guerra Grangeia
Faculty of Medical Sciences
University of Campinas
Campinas, Brazil
tguerra@unicamp.br

Marco Antonio de Carvalho Filho
Faculty of Medical Sciences
University Medical Centre Groningen
Groningen, The Netherlands
m.a.de.carvalho.filho@umcg.nl

2023 IEEE 19th International Conference on e-Science (e-Science) | 979-8-3503-2223-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/E-SCIENCE58273.2023.10254792

***Abstract*—As connected digital health data becomes more readily available, solutions are emerging to shorten the typical 17 years of latency in translating validated health knowledge into clinical practice. Learning Health Systems aims to achieve this goal. However, the proposed systems aim to address health data in a broad spectrum of data type variety. An open challenge is how to combine this variety around unification models. This work addresses a segment of this challenge by exploiting knowledge collected and built around Virtual Patients (VPs). VPs are a promising learning approach, providing interactive computer-based scenarios for solving clinical cases. Debate and resolution of clinical cases form the foundation of medical knowledge sharing and education. However, existing initiatives restrict their focus to a unidirectional method in which educators create these cases and learners play them. In this article, we show that we can expand the VP perspective toward a pivot model, which articulates learning and research initiatives, gathering together health knowledge. Our Jacinto platform and data space for sharing, learning, discussing, and researching clinical cases embodies this VP-centered approach. We present its effectiveness through a series of practical scenarios that explore and combine several knowledge pipelines.**

***Index Terms*—Electronic learning, Clinical diagnosis, Medical information systems, Virtual patient**

I. Introduction

Studies estimate around 17 years of latency to translate validated health knowledge into practice [1]. These observations motivated several propositions of Learning Health Systems [2]–[4]. They explore the widespread use of digital health data, from health centers to research institutes, to propose rapid learning infrastructures. Besides concerns of integration, dissemination, and access policies, one challenge is how to produce a unification model to congregate the diversity of health data. For example, one can relate data around a patient unification model, i.e., connect all information about lungs from a diversity of sources – clinical exams, ultrasounds, etc.

Our research started from a learning platform, electing the Virtual Patient model to design clinical cases to be solved by students. The importance of addressing a broad spectrum of clinical cases in medical formation fostered the development of learning approaches centered on solving virtual cases, besides real-life situations encountered by students in health facilities. The Virtual Patient (VP) refers to an approach involving computers to implement these cases.

Clinical cases are the basis of the development of cognitive schemas called illness scripts [5], which robustness is vital for diagnostic accuracy. Accurate case resolution and management is in the final stages of doctors' formation, as they articulate and integrate knowledge acquired during their entire course.

Existing platforms working with VPs focus on learning, mainly following the approach: educators write cases $\rightarrow$ learners play these cases. However, beyond a learning approach, clinical cases are also in mainstream medical communication and debates. This observation led us to produce the Semantic Virtual Patient model, a central contribution of this work, which unifies distinct health data sources, relating them around a pivotal patient model. It led our platform to new horizons of space for sharing, learning, discussing, and researching clinical cases.

The platform now runs on top of a Virtual Patient Data Space, where the Semantic Virtual Patient is also a pivotal model for healthcare data and evidence collection and documentation; explicit annotation and knowledge extraction from clinical cases and case resolutions; training and exploration of machine learning algorithms. Data Space here refers to an abstraction concerning several interrelated data sources with distinct integration levels. The platform does not define a Dataspace architecture as proposed by Halevy et al. [6], although this is a priority for the future.

One central platform player is a language that bridges the human perspective (authors and learners) of a Virtual Patient and the machine as a consumer of semantic information. In the author→language→machine direction, the language balances narrative freedom with machine-interpretable patient data by semantic annotations throughout free texts. In the author→language→learner direction, there is an expandable mapping mechanism of language elements into interface and interaction components family.

The platform proved to be a fertile field for: (1) exchange and co-construct knowledge about pathophysiology, history taking, clinical findings of diverse diseases, and point-of-care ultrasound; (2) explore and compare different strategies to teach clinical reasoning; (3) research diagnostic bias and the formation of illness scripts. Section IV details projects developed over Jacinto, emphasizing the distinctive aspects explored in the platform.

The remainder of the text has the following organization: Section II presents foundations and related work; Section III describes our platform and its contributions; Section IV details four practical projects over the Jacinto platform that illustrates the platform application and support our claims; Section V presents conclusions and future work.

II. Foundations and Related Work

This research addresses the Virtual Patient (VP) domain. The predominant perspective distinguishes VPs from other health learning approaches, which adopt aspects of patient simulation and clinical case resolution [7]. However, recent studies expand the VP notion to a broader spectrum [8], as approaches have the same focus and share common characteristics. From the broader VP perspective, authors distinguish initiatives according to technologies and competency they focus [9] plus other dimensions as game-based approach and authoring challenge [10].

We depart from the broader VP perspective in this study since distinct approaches carry complementary characteristics that underpin our proposal. Our focus here concerns the structure to represent patient data in the VP. Approaches vary in their representation according to the demand for machine-interpretable data. We selected four representative approaches for our analysis: document-based case scenarios, Virtual Patients (in the strict sense), virtual reality patients, and human patient simulators.

In Fig. 1, we confront two dimensions: patient data structuring and free-format narrative dependency. The more information is structured, the more machines can automatically interpret them. Highly structured data follow rigid schemas to represent each piece of data, e.g., patient temperature, heart rate, blood pressure, etc. It contrasts with the freedom narratives require, usually developed as a free text.

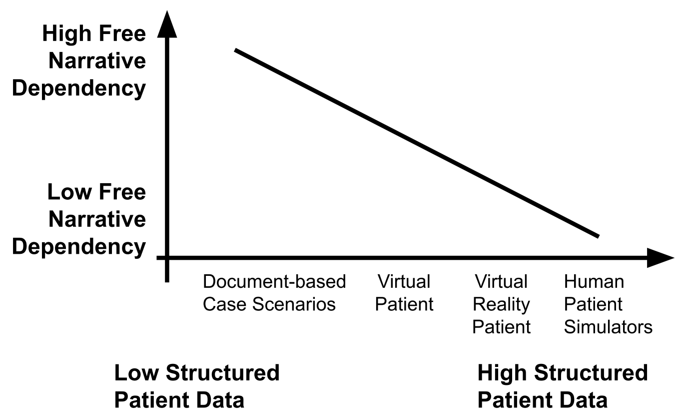

Fig. 1. Confronting narrative expressiveness and grade of structured information in different patient simulation approaches.

On one end, there are Document-based Case Scenarios. They are textual descriptions of cases presented as documents in paper or digital format. Each group discusses the case guided by a tutor. Since the document is an itinerary to direct the group work, there is a lot of flexibility and freedom in how to write it. Nevertheless, there are several initiatives to promote best practices and templates [11].

On the other end, there are Human Patient Simulators (HPS), based on automated manikins [12]. These simulators are as close as possible to a real human being. Its controlling software relies on highly structured information that establishes, for example, the heart rate and blood pressure of the manikin. The behavior and health condition of this manikin can pass through several states as part of a simulation scenario [13]. It is a graph where each node is a state of the manikin – i.e., a configuration of the simulator – and the transformations according to students' actions, e.g., one applies medication to the manikin to reduce the heart rate [14]. A simulation scenario is part of a narrative, which works in reactive mode. Educators conduct the main course of the narrative.

There are variants of patient simulation between these two poles. Some high-fidelity 3D simulations, based on virtual reality patients, have behavior close to the manikin but in the virtual context [15]. Here, besides the structured information that controls the "virtual manikin", there is space to introduce some free-text information that composes the narrative, e.g., the patient explaining his complaints.

What is defined as Virtual Patient (VP) [7] in a more strict sense has its roots in Document-based Case Scenarios, as they are mainly free-text narratives, along with images and videos. However, they introduce structured guidance in the form of a graph of states. Fig. 2 shows a simplified version of such a graph. Students start in the initial state (Patient Anamnesis) and, according to each action they take in the system, they change to a new state until they reach the final state.

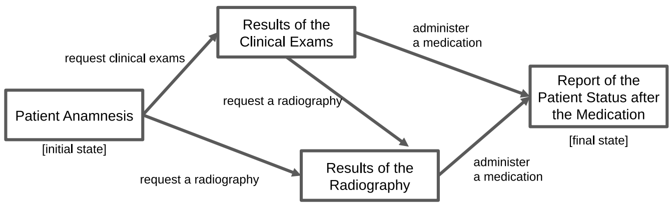

Fig. 2. Graph of states for a Virtual Patient.

Free texts, on the one hand, give the flexibility to produce scenarios at a lower cost as well as to describe more intricate narratives:

”The pain is continuous, and it does not get worse when I am breathing.”

”...my nose is stuffy, my throat is sore, and I am coughing, especially when I am in my bed, at night.”

On the other hand, more structured information can be interpreted by machines, expanding the possibilities of application: (i) data can subsidize richer simulations – e.g., a heart rate monitor; (ii) it becomes easier to find, reuse, and group cases and parts of cases – e.g., it becomes possible to query: cases in which the patient experienced shortness of breath; cases where the ECG was fundamental to diagnose a heart disease; (iii) data from Virtual Patients are a valuable source of health knowledge; when systematized in a Health Knowledge Repository can subsidize operations in the platform – e.g., support in the authoring process – and scientific research, as illustrated in Fig. 3.

Nevertheless, even systems with highly structured information, such as HPS systems, represent data in an internal format, limiting its application to a specific application.

Therefore, besides structured information, it is paramount to have an open standard to describe VP data. The MedBiquitous Virtual Patient (MVP) standard [16] is a standard defined by the MedBiquitous consortium and accredited ANSI [17]. It has several medical education community adopters such as CASUS, OpenLabyrinth, CAMPUS, and vpSim. MVP also subsidized the eVip – Electronic Virtual Patients project[1] that integrates several VP platforms in an open library containing 320 Virtual Patients.

MVP adopts XML and defines structures that distinguish the free-text narrative (VPDText) from structured data about the patient (PhysicalExam, DiagnosticTest, and Diagnosis). Studies over the eVip library showed two limitations addressed in this work.

[1]eVip project homepage: https://virtualpatients.eu

First, some initiatives tend to concentrate data in the free-text narrative (VPDText) [18]. We inspected some library samples and concluded two problems: (i) some platforms rely on free-text descriptions without structured patient data; (ii) there is no standard way to integrate patient data inside the narrative (VPDText) with structured data (PhysicalExam, DiagnosticTest, and Diagnosis). As will be described in Section III, our Versum language addresses this problem by integrating the narrative with structured data through semantic annotations.

Second, the XML basis of MVP enforces a rigid schema – with predefined interpretation – which is hard to expand. An example is the question and assessment items, not contemplated in the standard [19]. Each platform expanded MVP in its own way and question and assessment items from one platform are not recognized by the other.

There is a high diversity of approaches to implement Virtual Patients (VPs). Most of these initiatives address the resolution of clinical cases. However, there is a vast unexplored field in which students take the place of the author and learn by creating VPs. In the student-authored VP, the authors or co-authors of the Virtual Patient are the students. Generally, the students will create VPs based on cases encountered in their clinical rotation. They can record their findings in videos to enrich their cases. Besides improving clinical reasoning, this approach can enhance the student learning experience of a case [20]. Previous studies [20]–[22] did not utilize an specific tool for the students authoring – i.e., they utilized the same tool as the educators – and also did not guide them through the authoring process.

III. Jacinto Platform

Jacinto platform was born from observations of limitations in the initiatives to implement Virtual Patients:

1) the tradeoff between narrative freedom and structured patient data, as shown in Fig. 1;
2) lack of a systematic strategy to expand the platform to afford new interface and interface components;
3) support for student-authoring activities;
4) dynamic unfold of clinical cases supported by computational models.

Each observed limitation implied a corresponding fundamental characteristic of the Jacinto platform, listed in Table I.

Fig. 3 summarizes the chief cycles in the Jacinto platform. As the figure indicates, cycle 1 starts with the educator creating a Virtual Patient (VP) through the Authoring Environment (1a). It offers tools to produce multimedia cases and a toolbox with a broad spectrum of interactive components. The educator publishes this case into the Virtual Patient Data Space, sharing it with learners and educators, who will reuse or play it through a Player Environment (1b). The Progress Tracker module records all actions and decisions of each student for evaluation, tutoring, and scientific research. It stores records

TABLE I
FUNDAMENTAL CHARACTERISTICS OF JACINTO

language	a versatile language combining expressiveness for narrative and semantic annotations to extract machine-interpretable knowledge
mapping	a mapping mechanism to transform scripts into interactive VPs, based on an extensible library of web components
data space	a Virtual Patient Data Space architecture, in which learners also build and share VPs as part of a learning process
model	a Computation Model mechanism to drive clinical cases with dynamic unfold – e.g., a prognostic predictor according to learner decisions

in the Progress and Learning Data repository. Sections IV-A, IV-B, and IV-E present practical experiences in this cycle.

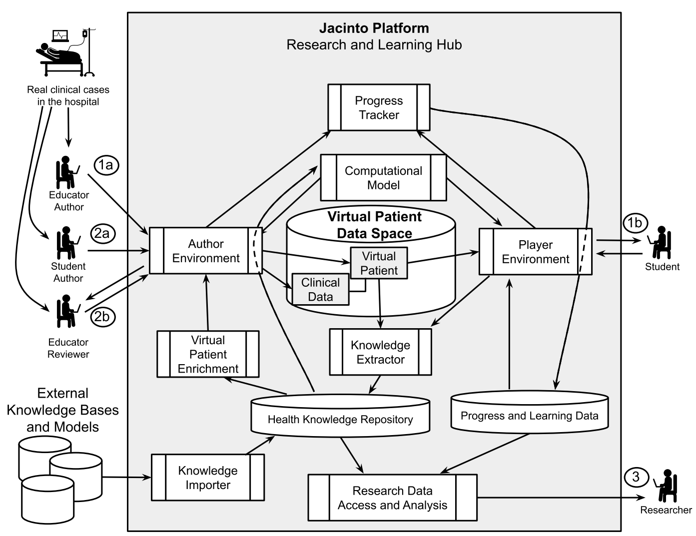

Fig. 3. Jacinto Platform as a Hub to consume and produce scientific knowledge.

In cycle 2, students assume the author role (2a). Departing from their experience as internal medicine residents, they produce a Patient Notebook for each case they follow, aggregating exams and data from these patients as their analysis and conclusions. Patient Notebooks follow the same VP model and the Virtual Patient Data Space stores them. Educators in this cycle advise and review students' work (2b). Section IV-D details this experience in the context of POCUS learning.

In cycles 1 and 2, authors follow templates having previously annotated elements or can annotate content during the authoring process. We term the first approach In Loco Semantics, wherein the context, defined by the location where the content gets inserted, allows for the inference of semantics through pre-annotated spots. The Knowledge Extractor module extracts annotated content transforming it into a Health Knowledge Repository. The content of this repository can, in turn, subsidize the Virtual Patient Enrichment process.

Data from the Health Knowledge Repository and the Progress and Learning Data feeds a module for Research Data Access and Analysis. Researchers from this project and external collaborators use this module to conduct scientific research concerning health and health learning (3). The Knowledge Importer module enriches and integrates the internal knowledge repository with External Knowledge Bases and Models. Section IV-D shows how annotated POCUS data subsidized training machine learning algorithms.

The Health Knowledge Repository also feeds a Computational Model module, which provides predictive and scoring capabilities based on statistical models for the Player and Author Environments. Section IV-E describes an experience based on a prognosis learning model (SAPS-3), which will expand to afford models computed by the platform.

A. Annotated Narrative

Any clinical case expressed as a Virtual Patient typically comes as a narrative [23]. It engages the participant in a challenge close to reality, where each decision will decide the Virtual Patient's fate. It also directs how the case unfolds along with time and choices, revealing new information close to the real-world pace.

Our Patient Notebooks approach also follows the VP model. In this scenario, students analyze the patients' cases and formulate their conclusions, producing the narrative.

Fig. 4 shows the Versum narrative language as a central element in the platform. Versum derives from Markdown and has structures inspired by Ink [24], a scripting language designed to develop interactive adventure games. Versum is semi-structured, interpretable by machines, and enables semantic annotations. As Markdown and Ink, authors can write narratives in Versum without any tailor-made tool. However, the platform also offers a full-fledged visual authoring tool. For detailed information about Versum see [25] and [26].

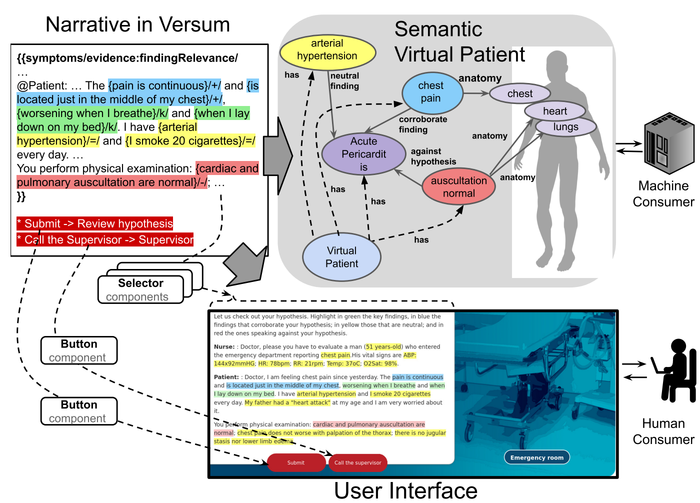

Fig. 4. Versum narrative scripting language as a bridge between dynamic clinical cases and semantic knowledge.

A Versum script has two faces. On the one hand, as the narrative unfolds in the Player Environment, a translator mechanism dynamically converts this script into interactive web pages populated by web components. Transformation templates guide the conversion process.

On the other hand, the language supports semantic annotation and the Semantic Extractor mechanism converts them into health knowledge with explicit semantics, as detailed in [27]. Students' textual productions during the resolution of a case are also subject to semantic annotations. As Section IV-C details, researchers can further annotate these texts for research purposes.

Semantic annotations are expensive to produce as they require extra effort besides narrative writing. We developed the In Loco Semantics approach that comprises two strategies to reduce the burden and motivate writers. First, templates drive the creation of the narrative backbone and each extra piece added to it. Semantic annotations populate these templates. An example, presented in Section IV-D, is a template to systematize ultrasound point of care (POCUS) videos. Since each video is from a specific human body part, the template conducts each insertion in a slot containing a description for humans and annotations for machines.

Second, semantic annotations are intrinsic to the method for building some case elements. The example presented in Fig. 4 shows a deliberate reflection strategy in a clinical case resolution, adopted in the Jacinto Bemelhor, described in Section IV-A. The system invites the player to reflect on a diagnostic hypothesis, informed in the previous screen, indicating which findings influenced the decision: green - key-finding; blue - corroborate finding; yellow - neutral finding; red - finding against the hypothesis. The case author must previously indicate the expected answer to each finding. The system uses it in the evaluation/feedback process, but it is also a rich source of semantic information relating signs and symptoms to diseases. In [27], there are details of how the annotated content becomes a knowledge graph relating signs/symptoms, diseases, and anatomical parts.

Besides human annotations, the platform also provides Natural Language Processing (NLP) methods to automatically annotate information of interest in texts, such as symptom identification, diagnostic hypotheses, and treatment decisions. A Named Entity Recognition (NER) module based on a neural large language model identifies and classifies these entities of interest within the text.

The Bio-Epidemiology-NER implementation [28] was used to perform NER of the texts of interest, as shown in Figure 6. The module enriches the recognized entities connecting them to the Medical Subject Headings - MeSH RDF ontology [29]. MeSH also enables identifying ignored entities. The module overlaps annotations, as detailed in the following subsection.

The platform organized all this knowledge around our pivotal model of Semantic Virtual Patient, illustrated in Figure 4 and introduced at [27]. It is an abstraction connecting the Virtual Patient to the knowledge network based on the Semantic Web.

B. Knowledge Pipelines

To combine knowledge, the model can integrate annotations in layers. Fig. 6 illustrates several integrated annotation layers, from different sources, around the same textual fragment. The Educator and NLP layers are real applications in the Jacinto Bemelhor context. We invited a researcher in the Evaluation of Illness Scripts project (Section IV-C) to annotate the same textual fragment for the sake of illustration, even though it is not a student's answer.

These annotations converge from different knowledge pipelines. The pipelines in the platform work around textual descriptions and semantics annotated or discovered in these texts. Consider these pipelines presented in Figure 5:

- Educators author Virtual Patients departing from pre-annotated templates (In Loco Semantics). For example, in the Jacinto Bemelhor project described in Section IV-A, templates have an area to report the findings (signs and symptoms) of a patient arriving at the hospital and their contribution to the diagnostics. As Figure 6 (layer Educator) shows, all the text in this area will automatically receive a "symptoms / findingRelevance" annotation.
  - During the authoring, the educator can manually introduce annotations that indicate those findings that contribute or not to the diagnostics. They will contribute to the student evaluation and feedback. These annotations are usually intrinsic to the case authoring process (In Loco Semantics) as the deliberate reflection strategy previously described.
- Researchers annotate textual answers produced by students to support research concerning new mechanisms to evaluate and give feedback. We detail an application of this pipeline in Section IV-C. In the Researcher layer of Figure 6, "chest pain" and "pain is continuous" received annotations indicating that they are part of the patient history.
- Students author Patient Notebooks departing from pre-annotated templates. Images, videos, comments, and conclusions introduced in this notebook fall into predefined areas related to semantic annotations (In Loco Semantics). We detail an application of this pipeline in Section IV-D.
- An NLP algorithm recognizes named entities, classifying and relating them to ontologies, as described in the previous section.

As shown in Fig. 6, the integration happens around the Virtual Patient as a pivot. This text is a fragment of a VP with Acute Pericarditis (semantically annotated in the document). Since annotations are in the context of the symptoms, it is possible to infer that continuous chest pain is a symptom of Acute Pericarditis. Common annotated textual elements allow inferring that the location of the pain is in the chest and that these are symptoms obtained through medical history. Moreover, being characterized as a continuous pain contributes to the specific diagnosis. Chest and pain here is not only a word but a concept in the MeSH ontology, enabling to make the semantics of this knowledge explicit.

This Semantic VP drives the integration of knowledge within a VP document but also across multiple VPs. In this way, it is possible, for example, to integrate knowledge (e.g.,

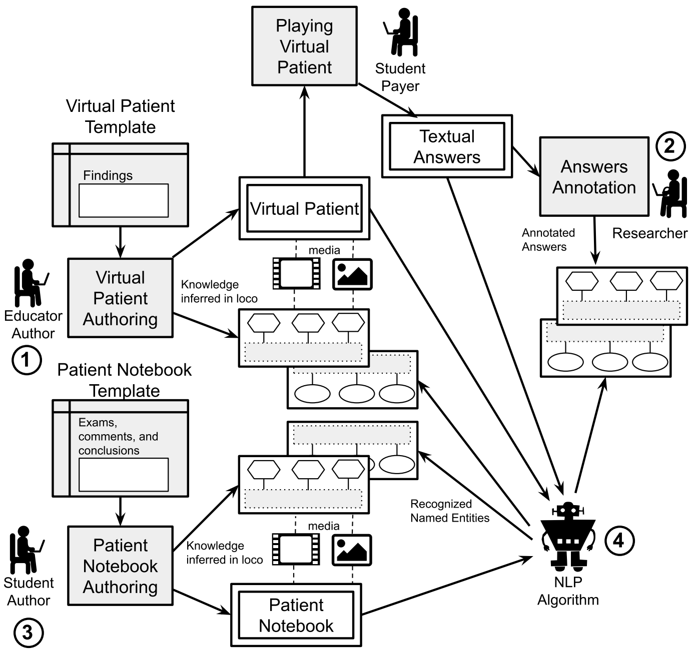

Fig. 5. Knowledge Pipelines.

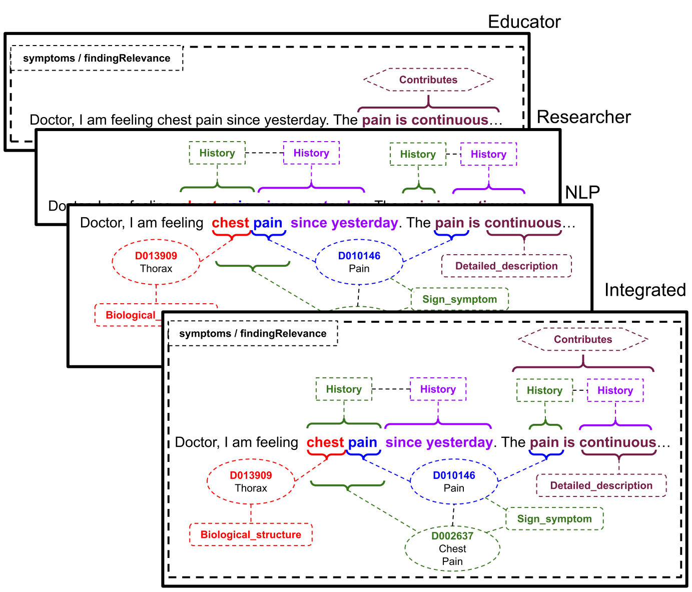

Fig. 6. Layers of semantic annotations (in loco, manual, and automatic) over the same text fragment.

symptoms) of several Acute Pericarditis VPs. It produces a Unified VP condensing the knowledge we have about Acute Pericarditis. It is also possible to focus on symptoms, e.g., what are the possible diagnostics for a patient with continuous chest pain?

C. Research Data Analysis and Access

The Jacinto Platform has an underlying data-gathering infrastructure that can be used as a source of scientific research. Much of the interactions in the system is with scientific information and educational purpose. Thus, gathering these interactions is essential to how the system can be used in medical education and understanding how physicians behave in simulated situations.

The data architecture currently in use in the platform uses two main approaches. First, a traditional relational database is used to store common user information, e.g., username, institution, role, etc. The second approach is using data streams to store user interaction with the player environment of the platform. The data streams are stored using Kafka and the document database MongoDB. This combination allows for the ingestion of huge amounts of data in a temporal fashion. The main strategy is to be able to track every user interaction in a session using the system and be able to replay this interaction later.

The data stream uses a messaging protocol based on MQTT. It uses message communication through the publish/subscribe abstraction, thus we have a source of data that can publish this data in a channel in which an interested consumer of data can receive it. The messaging protocol is essential for later data analysis, since the data stream stores fine-grained user interaction, a coarse-grained view of the data can be obtained by dividing the data into different channels, with different purposes, for different consumers.

The system also supports different pipelines of the same Virtual Patient for scientific research. For example, one experiment required two versions of the same clinical cases, only changing the order of the EKG presentation. Our data architecture allowed the medical education researchers to properly identify the students' responses, resolution time, and degree of confidence in the diagnosis for each version of the clinical cases. This allowed the researchers to derive insights into the student diagnostic process and even this impact on current medical practice.

IV. Research Showcase

This section presents a showcase of projects over the Jacinto platform, which we mentioned in the previous sections. They illustrate, by practical applications, that the characteristics listed in Table 1 (emphasized in bold) support sharing, learning, discussing, and researching clinical cases. They also implement and validate our proposition of knowledge pipelines, combined around a semantic Virtual Patient pivotal model. Table II summarizes key usage statistics of our showcase, which outlines in numbers the adherence to the platform.

Subsection IV-A starts with the original Jacinto experience on the platform, whose conception gave rise to language, mapping, and data space. Subsection IV-B presents a Team-Based Learning (TBL) approach, whose variety of narrative elements explore the flexibility of the language and the mapping. Subsection IV-C presents a research initiative to evaluate strudents answers based on the illness script framework. Subsection IV-D shows how students work as authors producing annotated UltraSound Point of Care (POCUS) media of patients in a shared data space. It emphasizes the role of the language embedding machine-interpretable annotations in

TABLE II
RESEARCH SHOWCASE IN NUMBERS

Project	Subsection	Numbers
Jacinto Bemelhor	IV-A	997 Virtual Patients
Team Based Learning	IV-B	4 tematic blocks 15 Virtual Patients 400 rounds of students
Evaluation of Illness Scripts	IV-C	12 Virtual Patients 8 open questions 723 students 1,430 textual answers 344 annotated texts 11,412 manual annotations
POCUS Learning	IV-D	48 doctors and students 1,870 patient notebooks
Prognostic Learning	IV-E	71 students

this media, further using them for scientific research. Subsection IV-E shows how students work as authors of Virtual Patients in an ICU to learn prognosis. A computational model supports predictions and evaluation of students' decisions.

A. *Jacinto Bemelhor (Phill Muchbetter)*

Jacinto is a project that started in 2013 in the School of Medical Sciences at Unicamp [30]. It achieved the development of more than a thousand clinical cases. In a five-year study [31], Jacinto served 462 students.

The platform layout and the case design followed the concepts of the Cognitive Load and Self-Determination theories [30], i.e., the cases were built from simple to complex, always taking into consideration the previous knowledge of students. At the same time, the platform and the cases optimized students' sense of autonomy, competence, and belonging.

In 2016, Jacinto started a new stage involving the system redesign, resulting in the platform presented in this work[2]. Schemas subsidized templates that guide the construction of Virtual Patients. These templates embed a long-term successful experience and also machine-interpretable annotations.

The project also expanded toward a platform that embraces narrative-based cases in other domains named Harena[3]. The Jacinto platform became the health branch of Harena. Detailed documentation is available at Harena Docs[4].

The next step of "Jacinto" is to allow multiple actors to co-construct the cases and exchange feedback in real time. We know that students construct their cognitive schemas by being exposed to real clinical cases. We also know that reflecting on these experiences and contrasting clinical cases with similar clinical findings is vital to enhance the quality of these representations. Thus, if we allow students to create their own cases and teachers and students to engage in reflective and dialogical conversations about these cases, we can boost script construction while guaranteeing their quality. This process demands for high levels of intrinsic motivation.

[2]Jacinto homepage: https://jacinto.harena.org
[3]Harena GitHub: https://github.com/harena-lab/
[4]Harena Docs: https://harena-lab.github.io/harena-docs/

Self-Determination theory states that developing intrinsic motivation depends on fulfilling their basic psychological needs: autonomy, competence and relatedness. The new "Jacinto" allows students to choose their learning trajectories and create their own cases. It will have adaptive learning algorithms to find the right level of difficulty based on students' prior performance to match the challenges with students' "zone of next development", which nurtures a sense of competence. Finally, by allowing conversations in real time, and creating a sense of identity, the new "Jacinto" helps to create a community, in which teachers and students are connected. So, "Jacinto" strategy aims to foster students' intrinsic motivation to learn.

B. *Team Based Learning*

The collaborative environment provided by Jacinto grounded its expansion to the Team-Based Learning (TBL) method. Educators can organize students in teams and control each step of this method. Students start with the individual readiness assurance test (iRAT), which is a series of multiple choice questions based on clinical cases. Next, the same questions are solved within the group (tRAT). After the iRAT and tRAT, the facilitator of the session gathers immediate feedback from the platform regarding students' performance. With this information, the facilitator may correct eventual misunderstandings that would jeopardize students' performance in the next step of the TBL: solving a series of clinical cases in groups. The platform allows students to work onsite and online. Each case design for teams fosters debate among students. The work is closed with a discussion with the facilitator, who has a panel to browse the cases solved by individuals and teams.

The TBL method required the expansion of interactive mechanisms, now designed to work with teams. A new management interface enabled grouping students in teams; to control the progressive unfolding of clinical cases; to follow students' progress online; and to navigate through their results.

C. *Evaluation of Illness Scripts*

Illness scripts serve as frameworks to store comprehensive knowledge about diseases in long-term memory. These scripts consist of three essential components: fault, enabling conditions, and consequences. The fault component identifies the underlying pathophysiological processes responsible for the disease. Enabling conditions encompass the baseline patient conditions, which may or may not act as risk factors for the development of specific diseases. Finally, the consequences include the clinical syndrome, including all pertinent clinical data and the typical tempo of progression. Similar to other cognitive psychology concepts, illness scripts are intangible constructs that cannot be observed or physically interacted with. We can only gain insights into them by analyzing the responses of physicians and medical students during specific experimental tasks.

One particular type of task that can be analyzed is referred to as a recall task. The other primary category of task is known as a recognition task. During recall tasks, individuals are prompted to retrieve information that is stored within their memory. Recognition tasks, on the other hand, involve the identification and acknowledgement of familiar data based on memory cues. One possibly effective approach to gain valuable insights into illness scripts involves the utilization of recall tasks. In this method, students are presented with the task of recollecting their knowledge about a particular disease, recording their response in writing, and subsequently interpreting their written answers. This method could provide a valuable means of assessing and understanding the depth of students' understanding regarding various diseases.

To interpret the texts, the first step is to employ a qualitative research concept called content analysis. The responses should be broken down into idea units, which will be classified by the annotator into categories relevant to illness script evaluation. An experiment was conducted using the Jacinto platform to accomplish this objective (not yet published). At the outset of their clinical years in medical school, students were requested to recollect their knowledge regarding two diseases: acute myocardial infarction and chronic obstructive pulmonary disease. The platform implemented a time constraint of five minutes, which could be enforced to restrict participants from continuing their writing. The responses generated were then disassembled into individual idea units and systematically annotated across several categories, encompassing medical dimensions such as pathophysiology, epidemiology, etiology, clinical features, and treatment. The platform facilitated researchers in effortlessly selecting idea units, grouping them together, and categorizing them accordingly. Moreover, it automatically calculated predetermined scores pertaining to organization, accuracy, and other relevant constructs associated with illness scripts. It is a powerful resource to analyze this type of data in a streamlined manner.

D. POCUS Learning

Ultrasound point of care (POCUS) uses a relatively old technology (ultrasound) differently. For decades, ultrasound was used as a formal exam performed or interpreted by a radiologist that is sometimes detached from the patient's clinical context. POCUS refers to the use of ultrasound by the physician responsible for taking care of the patient. It is a tool for improving the physical exam, clinical reasoning and the decision-making process. It is time-sensitive, directed by the problem, performed at the bedside and limited in the scope. Different specialties in medicine have applied it because it saves time, promotes more rational use of diagnostic resources, increases patient safety, and, more importantly, can help save lives in specific situations.

POCUS competency requires proper capacitation. Because of its wide applications, physicians need to learn how to use this technology to benefit their patients. It is of great importance to have trainee programs that allow the acquisition of competency for using the machines, acquiring good quality images with minimal criteria, interpreting the findings correctly, and integrating them into the clinical context of a specific case. To achieve those competencies, the physician, the resident or the medical student must be in a program that allows a learning environment with a varied range of cases, with supervised hands-on, autonomous practice and a collection of a portfolio of images for interpretation.

In a POCUS program for internal medicine residents (Hospital de Clínicas de Porto Alegre, HCPA), before using Jacinto's platform, the images acquired were in a drive linked with a case written on a sheet. The supervisor should review and give feedback. A very artisanal and rudimentary process. The platform came to structure and facilitate this part of the learning process. By putting their cases there, the physician/student can organize the clinical thinking, indicate which questions their exam can answer, choose the best video or image performed, describe the findings, interpret the results in the clinical context and make a decision about the case. It also allows for more structured feedback from the supervisor.

Since POCUS training relies heavily on lengthy hands-on training sessions and daily practice on image interpretation, even with the support provided by Jacinto's platform, a key challenge emerges: the insufficient volume and availability of expert faculty members that can assist trainees and provide feedback on their interpretation skills. Another challenge is the need of practitioners to maintain high levels of acuity in clinical practice, as both suboptimal images and image misinterpretation may critically affect patient management.

Recent advances in the field of artificial intelligence and machine/deep learning, namely in the medical arena, hold great promise towards tackling the above challenges, particularly with the development of automated image interpretation tools to virtually assist POCUS training and provide means for continuous medical education on this topic even under limited access to experts [32]. Potential use cases include automatic image quality assessment, image optimization assistance, or image labeling/interpretation allowing students' self-directed learning (in accordance with the Self-Determination theory).

Given its utility for POCUS training (as exemplified at HCPA POCUS internal medicine program) and the structured nature of the underlying data, the Jacinto's platform can help boost the development of these artificial intelligent systems by providing valuable curated data that can be fetched to train them. Interestingly, the co-existence of information regarding clinical presentation, imaging data (frequently targeting multiple organs) and diagnosis, besides expert feedback on all of them, further paves the way for AI-related research in fields like visual question answering, question-oriented case retrieval or automated decision support systems.

E. Prognostic Learning

Prognosis is the deduction of the probable course and outcome of a disease, especially important when it comes to a life-threatening one. To do so, patient and disease-related variables must be identified and put together to measure their impact on clinical outcome [33]. Traditionally, medical education has left prognosis out of the spotlight, which has been shared between diagnosis and treatment. Outstanding diagnostic tools and evidence-based treatment protocols balanced with an upcoming culture of patient-centered care have shifted physicians' attention toward prognosis. The understanding of the natural history of disease and disease-modifying treatments – as much as its adverse effects – allows a tailormade, value-oriented plan of care.

The Simplified Acute Physiology Score III (SAPS-3) [34] is a prognostic model based on patient variables at Intensive Care Unit (ICU) admission. Medical history, ongoing clinical status, laboratory findings, and reason for ICU admission are collected within 1 hour of patient admission aiming to predict in-hospital mortality. As a tool to measure ICU assistance quality, SAPS-3 is broadly applied in hospitals worldwide and its development was based on a global database. For its widespread validation and use, it was chosen as the prognostic model through which students would be exposed to prognosis learning.

Our work focused on utilizing Virtual Patients for medical prognosis learning in an ICU scenario. As presented in Section II, in the student-authored approach the learner takes the author role. The students would build a VP, constrained by a template following the parameters that SAPS 3 takes into consideration. After the creation of the VP, the student needs to predict the patient's survivability percentage chance. The student's prediction is then compared with the SAPS 3 prediction, giving instant feedback to the student about the VP's survivability. As far as we know, we are the first to utilize a prediction model to assist in the VP's authoring process, as well as guiding the student and providing instant feedback.

The use of a predictive model in dynamic interactions – see II – enables mainly two utilizations. The first one is to dynamically generate paths for the clinical case based on previous interactions with the VP, enriching the student's experience. The second is to utilize a predictive model in assisting the creation of VPs. Our work utilized a model based on the SAPS 3 predictive model [34] for assisting the students in the authoring process. Based on the information of the created VP (e.g., age, comorbidities), the model predicts the survivability of the Virtual Patient and compares it to the prediction given by the student.

V. Conclusions and Future Work

This paper presented our perspective of a Virtual Patient platform that expands the classic position of a tool, where educators write clinical cases and learners play them, towards a space for sharing, learning, discussing, and researching. In the theoretical context, it contributes with an innovative unification model (the Semantic Virtual Patient) and its related processes. In the practical context, our platform materializes this model in a collaborative environment.

Our Virtual Patient model became a pivot to integrating knowledge from different sources following complementary knowledge pipelines. Semantic annotations are captured and produced in distinct approaches, ranging from manual annotations, through annotations inferred by the context of their insertion or intrinsic to the authoring process (In Loco Semantics), to annotations produced automatically by an NLP module.

We showed how to organize knowledge around the Semantic Virtual Patient model and how it helps to integrate it within a VP document and across several VP documents. It subsidizes, for example, condensing knowledge about VP profiles, as a patient with Acute Pericarditis, or departing from a symptom, e.g., what are the possible diagnostics for a patient with continuous chest pain?

Through the research showcase, we presented how the fundamental characteristics of the platform supported our new perspective. The platform usage statistics in the showcase projects support our claims and the viability of our proposal.

The Virtual Patient Data Space perspective enabled going beyond educators as producers, including learners as authors – as a learning strategy and contributing to the knowledge base.

Future work includes advancing the Semantic Virtual Patient perspective and providing tools for exploratory analysis of Unified VPs. Additionally, there will be automatic assistance for VP authoring, utilizing the knowledge extracted and integrated into the platform. Improvements in our NER algorithm will involve integrating more ontologies and fine tuning it with annotations produced in our platform. For example, we are working on an approach to automatically recognize entities related to the illness script evaluation based on the manual annotations, described in Section IV-C.

Acknowledgments

This study was partially financed by CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) and by CNPq (Brazilian National Council for Scientific and Technological Development) grant numbers 428459/2018-8 and 400062/2023-2. This work was supported by Amazon Web Services, Inc. Financial support from the Foundation for Science and Technology (FCT, Portugal) in the scope of the project PTDC/EMD-EMD/1140/2020 and grant CEECIND/03064/2018 is gratefully acknowledged.

References

[1] Z. S. Morris, S. wooding, and J. Grant, "The answer is 17 years, what is the question: understanding time lags in translational research," *Journal of the Royal Society of Medicine*, vol. 104, p. 520, 2011. [Online]. Available: /pmc/articles/PMC3241518/ /pmc/articles/PMC3241518/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3241518/

[2] C. P. Friedman, A. K. Wong, and D. Blumenthal, "Policy: Achieving a nationwide learning health system," *Science Translational Medicine*, vol. 2, 11 2010. [Online]. Available: https://www.science.org/doi/10.1126/scitranslmed.3001456

[3] S. M. Greene, R. J. Reid, and E. B. Larson, "Implementing the learning health system: From concept to action," *Annals of Internal Medicine*, vol. 157, pp. 207–210, 2012.

[4] P. O. Vandvik and L. Brandt, "Future of evidence ecosystem series: Evidence ecosystems and learning health systems: why bother?" *Journal of Clinical Epidemiology*, vol. 123, pp. 166–170, 7 2020. [Online]. Available: http://www.jclinepi.com/article/S0895435620300779/fulltext http://www.jclinepi.com/article/S0895435620300779/abstract https://www.jclinepi.com/article/S0895-4356(20)30077-9/abstract

[5] H. G. Schmidt and R. M. Rikers, "How expertise develops in medicine: knowledge encapsulation and illness script formation," *Medical Education*, vol. 41, pp. 1133–1139, 12 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2923.2007.02915.x https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2923.2007.02915.x https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.2007.02915.x

[6] A. Halevy, M. Franklin, and D. Maier, "Principles of dataspace systems," *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '06*, pp. 1–9, 2006. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1142351.1142352

[7] D. A. Cook and M. M. Triola, "Virtual patients: a critical literature review and proposed next steps," *Medical Education*, vol. 43, pp. 303–311, 4 2009.

[8] N. P. A. Quail and J. G. Boyle, "Virtual patients in health professions education," pp. 25–35, 2019.

[9] A. A. Kononowicz, N. Zary, S. Edelbring, J. Corral, and I. Hege, "Virtual patients - what are we talking about? a framework to classify the meanings of the term in healthcare education," *BMC Medical Education*, vol. 15, pp. 1–7, 2 2015.

[10] T. B. Talbot, K. Sagae, B. John, and A. A. Rizzo, "Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations," *Int. Journal of Gaming and Computer-Mediated Simulations*, vol. 4, pp. 1–19, 1 2012.

[11] T. Chaplin, B. Thoma, A. Petrosoniak, K. Caners, T. McColl, C. Forristal, C. Dakin, J. F. Deshaies, E. Raymond-Dufresne, M. Fotheringham, D. Ha, N. Holm, J. Huffman, A. M. Lonergan, G. Mastoras, M. O'Brien, M. R. Paradis, N. Sowers, E. Stern, and A. K. Hall, "Simulation-based research in emergency medicine in canada: Priorities and perspectives," *Canadian Journal of Emergency Medicine*, vol. 22, pp. 103–111, 2019.

[12] L. A. Peteani, "Enhancing clinical practice and education with high-fidelity human patient simulators," *Nurse educator*, vol. 29, pp. 25–30, 2004.

[13] A. Takeuchi, T. Kobayashi, M. Hirose, T. Masuda, T. Sato, and N. Ikeda, "Arterial pulsation on a human patient simulator improved students' pulse assessment," *Journal of Biomedical Science and Engineering*, vol. 05, pp. 285–289, 2012.

[14] P. Andersen and K. Cox, "Simman 3g: Manikin-led simulation orientation," *Clinical Simulation In Nursing*, vol. 40, pp. 1–6, 3 2020.

[15] K. M. Freeman, S. F. Thompson, E. B. Allely, A. L. Sobel, S. A. Stansfield, and W. M. Pugh, "A virtual reality patient simulation system for teaching emergency response skills to u.s. navy medical providers," *Prehospital and Disaster Medicine*, vol. 16, pp. 3–8, 2001.

[16] V. Smothers, B. Azan, and R. Ellaway, "Ansi /medbiq vp.10.1- 2010 - medbiquitous virtual patient specifications and description document," p. 128, 2010.

[17] A. A. Kononowicz, A. J. Narracott, S. Manini, M. J. Bayley, P. V. Lawford, K. McCormack, and N. Zary, "A framework for different levels of integration of computational models into web-based virtual patients," *J Med Internet Res*, vol. 16, p. e23, 1 2014.

[18] N. Zary, I. Hege, J. Heid, L. Woodham, J. Donkers, and A. A. Kononowicz, "Enabling interoperability, accessibility and reusability of virtual patients across europe – design and implementation," pp. 826–830, 2009.

[19] A. A. Kononowicz and I. Hege, "Virtual patients as a practical realisation of the e-learning idea in medicine," 4 2010.

[20] C. Trace, S. Baillie, and N. Short, "Development and preliminary evaluation of student-authored electronic cases," *Journal of Veterinary Medical Education*, vol. 39, pp. 368–374, 1 2012.

[21] M. Imison, M. Imison, and C. Hughes, "The virtual patient project: Using low fidelity, student generated online..." vol. 2008, 2008, pp. 441–445.

[22] I. Hege, A. A. Kononowicz, and M. Adler, "A clinical reasoning tool for virtual patients: Design-based research study," *JMIR Medical Education*, vol. 3, pp. 1–11, 11 2017.

[23] R. Ellaway, T. Poulton, U. Fors, J. McGee, and S. Albright, "Building a virtual patient commons," *Medical Teacher*, vol. 30, pp. 170–174, 2008. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01421590701874074

[24] Inkle, "Writing with ink," p. 2021. [Online]. Available: https://github.com/inkle/ink/blob/master/Documentation/WritingWithInk.md

[25] M. F. de Menezes Mota, F. L. Pantoja, M. S. Mota, T. de Araujo Guerra Grangeia, M. A. de Carvalho Filho, and A. Santanchè, "Analytical design of clinical cases for educational games," vol. 11863 LNCS. Springer, 11 2019, pp. 353–365.

[26] A. Santanchè, "Versum narrative scripting language," 2022. [Online]. Available: https://harena-lab.github.io/harena-docs/versum/

[27] F. Pantoja., M. Filho., and A. Santanchè., "Harena semantics: A framework to support semantic annotation in citizen science systems." SciTePress, 2022, pp. 336–343.

[28] S. Raza, D. J. Reji, F. Shajan, and S. R. Bashir, "Large-scale application of named entity recognition to biomedicine and epidemiology," *PLOS Digital Health*, vol. 1, p. e0000152, 2022.

[29] B. Bushman, D. Anderson, and G. Fu, "Transforming the medical subject headings into linked data: Creating the authorized version of mesh in rdf," *https://doi.org/10.1080/19386389.2015.1099967*, vol. 15, pp. 157–176, 10 2016. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/19386389.2015.1099967

[30] T. D. A. G. Grangeia, B. D. Jorge, D. Franci, T. M. Santos, M. S. V. Setubal, M. Schweller, and M. A. D. Carvalho-Filho, "Cognitive load and self-determination theories applied to e-learning: Impact on students' participation and academic performance," *PLOS ONE*, vol. 11, p. e0152462, 3 2016.

[31] T. de Araujo Guerra Grangeia, B. de Jorge, D. Cecílio-Fernandes, R. A. Tio, and M. A. de Carvalho-Filho, "Learn+fun! social media and gamification sum up to foster a community of practice during an emergency medicine rotation," *Health Professions Education*, vol. 5, pp. 321–335, 12 2019.

[32] H. Shokoohi, M. A. Lesaux, Y. H. Roohani, A. Liteplo, C. Huang, and M. Blaivas, "Enhanced point-of-care ultrasound applications by integrating automated feature-learning systems using deep learning," *Journal of ultrasound in medicine*, vol. 38, pp. 1887–1897, 7 2019.

[33] L. C. Yourman, S. J. Lee, M. A. Schonberg, E. W. Widera, and A. K. Smith, "Prognostic indices for older adults: A systematic review," *JAMA*, vol. 307, pp. 182–192, 1 2012.

[34] R. P. Moreno, P. G. Metnitz, E. Almeida, B. Jordan, P. Bauer, R. A. Campos, G. Iapichino, D. Edbrooke, M. Capuzzo, and J. R. L. Gall, "Saps 3 - from evaluation of the patient to evaluation of the intensive care unit. part 2: Development of a prognostic model for hospital mortality at icu admission," *Intensive Care Medicine*, vol. 31, p. 1355, 2005.