

# Linking Knowledge for Simulation Learning

Irene Celino and Daniele Dell'Aglio

CEFRIEL – ICT Institute, Politecnico of Milano  
via Fucini 2, 20133 Milano, Italy  
{irene.celino,daniele.dellaglio}@cefriel.it

**Abstract.** Simulation Learning is a frequent practice to conduct near-real, immersive and engaging training sessions. AI Planning and Scheduling systems are used to automatically create and supervise learning sessions; to this end, they need to manage a large amount of knowledge about the simulated situation, the learning objectives, the participants' behaviour, etc.

In this paper, we explain how Linked Data and Semantic Web technologies can help the creation and management of knowledge bases for Simulation Learning. We also present our experience in building such a knowledge base in the context of Crisis Management Training.

**Keywords:** Linked Data, Simulation Learning, Planning, Provenance, Semantic Web.

## 1 Introduction

Traditional research on Semantic Web in e-learning [23,24] are aimed at promoting interoperability between training systems, thus usually the core investigation targets are standards and schemata to describe learning objects [15,5].

Our research is focused on a different kind of e-learning system, i.e. Simulation Training to improve soft skills [1]. In this context, not only it is needed to describe learning objects, but also to fully plan simulation sessions; those sessions should be interactive and engaging to challenge the trainees to improve their skills. Simulation Learning systems generally re-create a near-real environment for training sessions, in which learners are subject to stimuli: they have to learn how to deal with the simulated situation and how to react to it.. Such simulations need to be effective and engaging, so that the learners do not simply memorise notions about the specific matter, question or theme, but they actively and permanently acquire skills, practice and knowledge.

The scenario production is therefore the core and critical activity when building a Simulation Learning system. Knowledge technologies are needed to model and manage all the required information, often generated and managed by different and independent sources: scenario descriptions, events and stimuli for the trainees, storyboards for the learning sessions, multimedia assets, supporting documents and guidelines, trainees description and behaviour/decisions, learning session monitoring, etc. Such a wealth of information makes the Simulation Learning a knowledge-intensive context, which requires smart solutions.

We decided to adopt Linked Data and Semantic Web technologies to address the requirements of Simulation Learning. The knowledge diversity and scale calls for a solution which provides interlinking between different datasets while preserving possibly independent information sources; moreover, the knowledge coherence and consistency must be assured to guarantee the significance, meaningfulness and correctness of simulation scenarios and storyboards presented to trainees.

In this paper, we present our current effort in exploiting Linked Data and Semantic Web technologies to build a Knowledge Base for a Simulation Learning environment. We explain why we believe that the selected technologies not only offer a suitable means to knowledge representation and management, but they are specifically required to address the challenges of such an environment.

Section 2 introduces the basic concepts of Simulation Learning systems and a concrete scenario in Crisis Management Training; Section 3 details our exploration in the use of Linked Data and Semantic Web to build a Simulation Learning Knowledge Base illustrating the gained benefits; Section 4 specifies our modelling choices, while Section 5 suggests that such modelling could benefit from provenance tracking; finally, Section 6 concludes the paper.

## 2 Simulation Learning

Learning should be relevant to people’s workplace and lives: learning content should be truly understood, remembered and applied to actual practices. Only in this way, by actively engaging participants in experiential training, learners can apply their knowledge and learn the best practices [1]; more and more often, indeed, it is not enough to read information and listen to a frontal lecture.

In this section, we introduce the theme of Simulation Learning for Decision-making, we draw a generic architecture of a system to support Simulation Learning and we describe a concrete scenario that we will use throughout the paper to exemplify our approach.

### 2.1 Simulation for Decision-Making

Training plays an important function in the preparation of professional practitioners. Currently, there are two main modalities for such training: table-top exercises and real-world simulations. *Table-top exercises* are low cost and can be easily and frequently organised. However, they cannot create a believable atmosphere of stress and confusion, which is prevailing in real-life situations and is crucial to the training of timely and effective decision making. On the other hand, training through *simulation exercises* on the field can be very effective [6], but it is considerably more expensive, it can require specialist equipment and it can be difficult to organise.

Simulation exercises require an Exercise Director (or *trainer*) who plays a key role in every form of exercise: the trainer has access to the whole exercise programme, ensures that it proceeds according to a plan, often feeds information

to the “players” (the *trainees*) to let them make informed decisions in response (verbally or by written messages). Sometimes information fed to the trainees is timed in advance at pre-set intervals, regardless of the previous responses. However, flexibility allows a trainer to use judgement and experience in timing the inputs: his/her role should be aimed to facilitate rather than orchestrate the exercise, thus intervention should be minimal and trainees should be given time to recognise and correct problems. Nevertheless, usually it is up to the trainer to decide, for example, how much advice to give to trainees.

## 2.2 Architecture of a Simulation Learning System

The architecture of a Simulation Learning System is depicted in Figure 1. In the picture, we can identify the involved actors, which are the *trainees* – the learning participants engaged in the simulation – and the *trainer* – who activates the exercise and monitors the progress of actions during the training session.

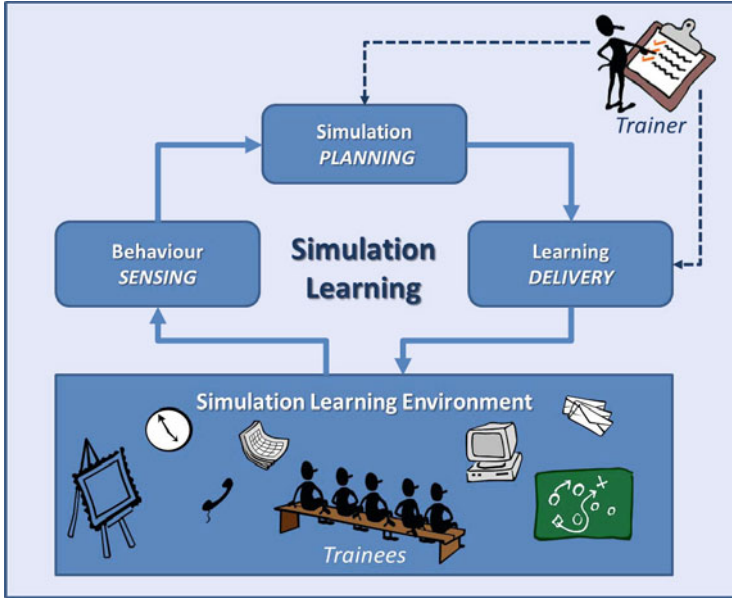
The figure also shows the four main modules of such an architecture, the first three following the usual AI sense-think-act cycle:

- *Behaviour Sensing*: this module is aimed to create and update a model of each trainee from sensors information (e.g. heart rate, blood pressure, respiration); the model represents trainee’s future and actual behaviour and provides indications on how to personalise the training path.
- *Simulation Planning*: this module is aimed to create and simulate a training scenario and its evolution, by combining the information in the behavioural model with knowledge about the learning scenarios; the output of this module is the actual simulation storyboard presented to the trainees.
- *Learning Delivery*: this module is aimed to effectively represent the simulation storyboard in the learning environment, including the rendering of audio-video inputs or Non-Player Characters (NPC, cf. Section 4.3).
- *Simulation Learning Environment*: this is the “place” where the training is conducted; the location can be a physical room or a virtual environment where the trainees interact and receive stimuli during a learning session.

The core of such system is therefore the Simulation Planning module, which contains the basic engine for creating active exercises for classes of trainees. The module is responsible for deciding which stimuli are sent to trainees and how they should be coordinated to create a meaningful and effective lesson plan. In broad terms, it is responsible for allocating over time the set of lesson stimuli indexed according to differences in presentation media, emotional characterization, personalization needs, etc.

## 2.3 Crisis Management Training Scenario

There is increasing recognition for the need to train non-technical skills like control and decision making for Crisis Management in national emergencies, high-reliability industries, as well as in industrial workplaces [20,22]. In the happening



**Fig. 1.** High-level architecture of a Simulation Learning System (from the classical sense-think-act cycle of AI)

of a catastrophic event, it is human behaviour – and often human behaviour alone – that determines the speed and efficacy of the crisis management effects [17].

The Pandora project<sup>1</sup> aims to provide a framework to bridge the gap between table-top exercises and real-world simulation exercises for Crisis Management, providing a near-real training environment at affordable cost. Its training system captures the good practice tenets of experiential learning but with greater efficiency and focuses on real, rather than abstract learning environments. The effective use of integrated ICT reduces the high dependence upon the trainer that is currently required to deliver exercises. Moreover, the Pandora framework supports the measurement and performance assessment of Crisis Managers, the key decision makers participating in a training exercise event as trainees.

As such, Pandora is developing an enabling technology to simulate believable dynamic elements of an entire disaster environment by emulating a crisis room (the Simulation Learning Environment). In this context, we are developing a Knowledge Base that makes use of Linked Data and Semantic Web technologies to model and interlink the pieces of data needed in the training simulation sessions. In the rest of the paper, we will use the Crisis Management scenario to exemplify our approach.

<sup>1</sup> Cf. <http://www.pandoraproject.eu/>

### 3 Our Simulation Learning Linked Knowledge Base

Within a Simulation Learning system, knowledge exchange plays a central role. In this section we give some details about the Simulation Planning module, focusing on the requirements, design and implementation principles of its Knowledge Base. All the technical details are related to the choices made in the Pandora framework.

#### 3.1 Knowledge Required to Plan a Simulation

To formalize the lesson plan, it is natural to choose a basic representation from timeline-based planning [10]. A plan is represented as a set of events having a temporal duration, distributed over a time horizon and indexed according to distinct features which should be planned for. This set of events is organized inside a data structure called *Event Network*, very common in current state of the art planning technology. The Event Network is a temporal plan of multimedia communicative acts toward trainees (e.g., e-mail messages, video news from an emergency location, etc.).

The Event Network can be generated by a *Simulation Planner*. This planner compiles static information into the Event Network, and then adapts the events configuration according to the actions of the trainees, thus simulating different courses of action of the world. The planner can be adapted from a generic AI Timeline-based Planning and Scheduling module [10].

The core information item elaborated by a Simulation Planner is the so-called *synchronization*. Synchronizations are the causal rules that regulate the transitions between values on the same planning feature and the synchronization of values among different planning features. In the Crisis Management scenario, synchronizations are used to influence the Crisis Managers' decisions, e.g. to generate changes in the emergency conditions.

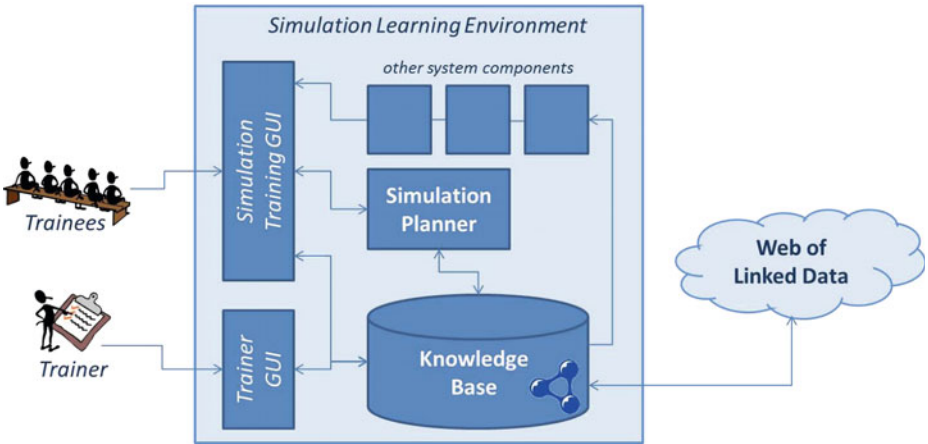
When adopting Planning and Scheduling technologies to simulate a scenario, it is worth highlighting how a great effort and amount of time is necessary to understand the problem, capturing all its specificity, and to create a model of the relevant aspects of the domains and the problem [9]. This consideration suggests, on the one hand, the need for identifying commonalities and similarities among the different domains and problems to operate in a more systematic way and, on the other hand, the opportunity to exploit Semantic Web technologies to ease and support the knowledge modelling task.

For those reasons, we have built a *Knowledge Base* with Linked Data and Semantic Web technologies. This KB is a central component in the Simulation Learning system, responsible for collecting and maintaining the "knowledge" about scenarios and training sessions. As such, the KB is the core information source for the simulation: it contains all the knowledge required by the Simulation Planner to "orchestrate" the events during the training sessions. All the causality in a simulation domain is modelled and stored in the KB; this knowledge is then converted by the Simulation Planner into the suitable data structures to synthesize the Event Network configurations for the lesson plan goals.

### 3.2 Requirements for the Knowledge Base

The Knowledge Base [8] was carefully designed to fulfil a pressing requirement: containing and managing all the knowledge needed to model and run the simulation scenarios, the training events, the trainees’ behaviour, the time sequence, and so on.

To fulfil such a requirement, the KB must reuse pre-existing information (e.g., in the Crisis Management scenario, training procedures, emergency management guidelines) and, in the meantime, it must allow for customization and diversification of training knowledge (e.g., emergency policies and legislation change from country to country). Furthermore, since most of the related information can be pre-existing in a variety of formats, the KB must be able to gather information from heterogeneous sources (e.g., location data from geographic datasets, audio and video inputs from multimedia archives, participants profiles) and to synthesize and interlink such knowledge into a coherent base.



**Fig. 2.** Role of the Knowledge Base in a Simulation Learning Environment

The role of the KB in the Simulation Learning Environment and its interactions with other components is depicted in Figure 2:

- The KB is “initialized” by the trainer who *models the simulation scenarios* and the training path alternative options;
- It is accessed by the Simulation Planner that needs to understand what “*events*” should be triggered and presented to the trainees during the learning sessions;
- It is also accessed by other system components that need to get/give *information about the training session* and the knowledge exchanged during or after its delivery (cf. Section 4);

- It is used to record the events and decisions taken during each training session, in order to enable the semi-automatically *creation of an individual trainee debriefing report* at the end of the training session.

To cope with such challenges, we adopted Linked Data and Semantic Web technologies for the design and development of our Knowledge Base.

### 3.3 Benefits from the Adoption of Linked Data

The choice of Linked Data and Semantic Web technologies in our KB is motivated by the need for an easy access, (re)use and integration of data and knowledge [14].

The *ease of access* to the KB is implicit in the use of Web technologies, which represent a mature and established technology stack. Following the Linked Data principles [2], we provide a standard access means to the data and knowledge stored in the KB. Moreover, Linked Data and Semantic Web facilitate and enable an entity-centric design of Web APIs: in our implementation, on top of the KB, we have developed a RESTful service<sup>2</sup> with specific methods to get details about certain entities on the basis of the concepts (entity types) defined in the KB ontologies and models (cf. Section 4). The RESTful service is also employed to abstract from the physical location of data, as explained further on.

The *reuse of pre-existing datasets* is also enabled by our technological choice. Several useful data sources are already present on the Web of Data and, thus, immediately exploitable by the KB. For example, in the Crisis Management scenario, environment characteristics of crisis settings are retrieved from GeoNames<sup>3</sup>, the geographical database containing over 10 million geographical names, 7.5 million unique features, 2.8 million populated places and 5.5 million alternate names. For example, a scenario about a river flood or a earthquake benefits from the retrieval of localized information from GeoNames. As a pragmatic solution, we are “caching” the relevant features from GeoNames locally to the KB. However, the reuse of GeoNames URIs constitutes a link to the remote dataset and allows for further knowledge retrieval. Specifically for the UK, geographic data come also from OrdnanceSurvey<sup>4</sup>, a government mapping initiative that also releases open data. In the same way, we can connect the KB to other knowledge bases like Freebase<sup>5</sup> or DBpedia<sup>6</sup> [4] to get information on a number of general-purpose topics and entities. The linkage to the latter sources is still in progress.

But this re-usability benefit applies also to the knowledge explicitly modelled for domain-specific learning scenarios: the choice of RDF to encode the data and of RDFS/OWL to model their structure pays, since those data are partially published on the open Web, thus enriching the Web of Linked Data

---

<sup>2</sup> Cf. <http://pandoratest01.xlab.si:8080/pandora-ckb/>

<sup>3</sup> Cf. <http://www.geonames.org/>

<sup>4</sup> Cf. <http://www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html>

<sup>5</sup> Cf. <http://freebase.com/>

<sup>6</sup> Cf. <http://dbpedia.org/>

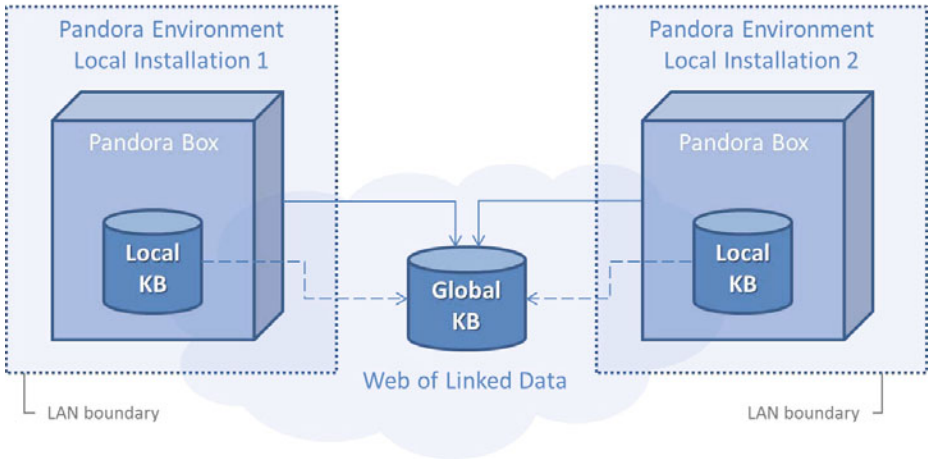
and becoming available for other Simulation Learning systems or for different tools. To this end, in our Crisis Management scenario, we decided to store the schemata and data generated by Pandora components natively as RDF triples in the KB; the knowledge coming from pre-existing sources in different formats (e.g., taxonomies, spreadsheets, guidelines) have been converted – manually or, whenever possible, semi-automatically – to a structured RDF format. The benefits of this approach are: the general Crisis Management knowledge is available to the whole community; the simulation scenarios can be reused by any installation of the training system; the further enhancements and extensions of the core knowledge are immediately “reflected” in all systems that make use of our KB.

The *ease of integration* comes from the native interlinking capability of Linked Data technologies. RDF provides the basic mechanism to specify the existence and meaning of connections between items through RDF links [3]. In other words, through the adoption of RDF, we not only give a structure to the data stored in the KB, but we also interlink the entities described by such data. Moreover, the links drawn between knowledge items are typed, thus conveying the “semantics” of such relationships and enabling the inference of additional knowledge. The information sources of the KB can be maintained and evolve over time in an independent way, but, in the meantime, can be connected via the Linked Data lightweight integration means.

The KB contains different (although interlinked) datasets, which also require diverse confidentiality/security levels for management and access. To this end, the KB can be designed as a set of federated RDF stores. The shared knowledge (e.g. general Crisis Management information, basic scenarios) could be “centralised”, to let all training system instances access and use it, while the installation-specific knowledge (e.g., detailed or customized scenarios, trainees information, personalizations) could be managed in a local triple store, not accessible from outside the system (see Figure 3). The RESTful service on top of the KB, as explained earlier, can provide a uniform access to the KB and hide to other accessing components the existence of the various “realms” of distinct Linked Data sources.

In the Pandora project, the choice was made to have the KB as a unique triple store with all the integrated knowledge. In fact, while the federated design was recognized as an interesting and meaningful solution, the distinction between “local” and “global” knowledge within the KB was very hard if not impossible to define. Indeed, some intrinsically non-confidential knowledge like the simulation scenario descriptions can represent a valuable information asset for a training company, which can be unwilling to make this knowledge freely available as Linked Open Data to competitors. Conversely, while the possibility to introduce customizations and personalization in the simulation scenarios is a very appreciated feature, making modifications to a global copy of that kind of knowledge would impact on all Pandora installation using that knowledge. Summing up, while Linked Data technologies are acknowledged as a good technological choice, Linked Open Data are not always applicable because of business or pragmatic reasons.





**Fig. 3.** The KB as federation of different triple stores to preserve security and confidentiality while benefiting from interlinking

Finally, the adoption of Semantic Web technologies in the form of ontologies and rules provides a further gain, since we can exploit *reasoning and inference* for knowledge creation and consistency checking, as explained in next section.

## 4 Modelling and Retrieval in Our Knowledge Base

As previously mentioned, our Knowledge Base manages several different and interlinked types of information. In this section, we introduce three “families” of data included in the KB and explain their modelling choices. We also illustrate their use in the Crisis Management training scenario within the Pandora Integrated Environment.

### 4.1 User Modelling

As introduced in Section 2.2, a Behaviour Sensing module is devoted to the “detection” of trainees’ performance in order to create individual models that help in tailoring the learning strategy of each participant to the simulation. Prior to the training session, dedicated psychological tests and physiological assessment at rest (e.g., through a Holter that measures the heart rate activity at rest) are used to measure some relevant variables (like personality traits, leadership style, background experience, self-efficacy, stress and anxiety). Those variables are then updated during the training session, through self-assessment measurements (i.e., asking the trainee about his performance) or through the elaboration of the raw data recorded by the sensors.

Those data about trainees’ behaviour are stored and updated in our KB, as instances of ontology concepts that represent the “affective factors” that influence the decision-making of the trainees. Due to the sensitivity of such information, the individual performances of the trainees are modelled in RDF and stored in the “local” triple store (cf. Figure 3) for apparent privacy reasons. We are also investigating the possibility to exploit Named Graphs [7] for access control: if the training session recordings are “stored” in the KB as separated named graphs, a named graph-aware access control component could grant admission to the allowed users (e.g., the trainer) and could deny the access of the malicious or occasional users (e.g., the other trainees).

In the specific scenario of the Pandora Integrated Environment, the learning sessions are targeted to the training of Crisis Managers. Therefore, the KB stores and manages also a set of specific information about them.

The Crisis Managers are the so-called Gold Commanders, who are responsible for the strategic development of responses to crisis situations. The trainee group is usually composed of the representatives of the “command team”, i.e. the core agencies involved in the strategic Crisis Management (e.g., police, local authority, fire brigade, ambulance); sometimes, other trainees can come from other utility companies (e.g. electricity, road transportation, environmental agency).

In our KB, therefore, we modelled the basic knowledge about those Gold Commanders by creating classes to represent the different trainees typologies. Those classes are “instantiated” per each training session, by adding the individual trainees to the KB. This lets the system record the training of each participant in relation to his/her role in the simulation; this knowledge is very precious for both the debriefing phase – when the trainer summarizes the performance results of each trainee (see also below) – and for a general analysis and mining of the achieved objectives and learning needs of the different agencies.

The initial version of the user modelling is part of the Pandora Ontology<sup>7</sup>.

## 4.2 Training Simulation Modelling

The core module of the simulation learning system is the Simulation Planning (cf. Section 2.2). Our KB therefore must be able to manage the knowledge required for the planning, in terms of the basic entities used by AI Planning Applications based on Timeline Representations.

In literature, several attempts tried to formalize the semantics of planners [19,12]. However, those approaches, on the one hand, tried to specify a generic planning ontology and, on the other hand, were specifically tailored to some application domains.

Building on their experience, we decided to make our own formalization to encompass the family of techniques known under the name of Timeline-based Planning and Scheduling. In fact, current AI planning literature shows that timeline-based planning can be an effective alternative to classical planning for complex domains which require the use of both temporal reasoning and scheduling features [10]. Moreover, our modelling aims to become the foundation for the

<sup>7</sup> Cf. <http://swa.cefriel.it/ontologies/pandora>

investigation on the interplay between Semantic Web Technologies and Planning and Scheduling research [8]; Semantic Web knowledge bases, in fact, can represent a good alternative to the current domain modelling in the planning area, which encompasses a multitude of custom and not interoperable languages.

Our modelling is formalized in a Timeline-based Planning Ontology<sup>8</sup>. As in classical Control Theory, the planning problem is modelled by identifying a set of relevant features (called *components*) which are the primitive entities for knowledge modelling. Components represent logical or physical subsystems whose properties may vary in time; in the simulation learning, components are either trainees behavioural traits or learning scenario variables. Their *temporal evolutions* is controlled by the planner to obtain a desired *behaviour*. Therefore, our ontology includes a set of *time functions* that describe the evolution over temporal intervals. The evolution is modelled by *events* happening on modelled components. To this end, a set of *planning rules* (or synchronizations) specifies what events can be triggered to modify these evolutions. The task of the Simulation Planner is to find a sequence of events that brings the system entities into a desired final state.

The core concept of the Timeline-based Planning Ontology is therefore the planning rule: each rule puts in relation a “reference” event – which is the potential *cause* of some phenomena in the simulation – with a “target” event – which is the possible *consequence* –, under a set of conditions called *rule relations*. We modelled such conditions as SPARQL FILTER or LET clauses<sup>9</sup>; therefore, we reused the modelling of such clauses and functions included in the SPIN Modeling Vocabulary [16] and extended it with regards to temporal conditions.

At learning *design time* – i.e. prior to the simulation sessions –, the trainer has to model the possible training scenarios, by instantiating in the KB the ontology concepts, in particular the planning rules and the related events. The choice of Linked Data and Semantic Web technologies for our modelling is not only useful for reusing and exploiting pre-existing knowledge. In this case, we can also exploit the semantics of such ontology for the *consistency checking* of the simulation scenarios: by automatic means, we can check if all the planning rules are satisfiable, if they represent possible “states” of the world simulated during the sessions, if all the events can happen under opportune conditions, and so on.

At *run-time* – i.e. during the simulation learning sessions –, all the events and decisions taken by the trainees during their learning are recorded in the KB. The KB is therefore used by the Simulation Planner to create and update the simulation plan. SPARQL-based querying is used to perform the knowledge retrieval required in this step: based on the actual recorded events, only the admissible planning rules are returned to let the planner decide what events to trigger.

---

<sup>8</sup> Cf. <http://swa.cefriel.it/ontologies/tplanning>

<sup>9</sup> The SPARQL LET clause is defined in some implementations, like the Jena Semantic Web Framework <http://openjena.org/>

After the learning session, at *debriefing time*, the recording of trainees’ behaviour and decision-taking is exploited to summarize the session progress. Also in this case, SPARQL-based querying on the KB is exploited to retrieve all the events and situations that involved each trainee; this knowledge is immediately at disposal of the trainer to produce a debriefing report for each participant and can be used to highlight personal performance, achieved training goals and attention points for improvement or further training.

### 4.3 Asset Modelling

The Learning Delivery module (cf. Figure 1) takes as input the simulation plan and “execute” it by sending the opportune stimuli to the trainee. To do this, it needs to recreate the actual simulation conditions, by pretending a near-real situation. For example, in the Crisis Management training scenario, the participants must be solicited by phone calls, mail, news, videos, etc. that give them updates on the evolution of the emergency. To this end, the Learning Delivery module manages two types of “learning objects” that are described in the KB.

The first type of simulation objects consists in *audio and video assets*, which give information to the trainees about what happens outside the simulation room. In the Pandora scenario, those assets are pre-canned recording of simulated video news or audio inputs – like phone calls from the crisis setting – which are used to put pressure on the trainees and, in the meantime, to give them further inputs on which they must base their decisions. To model such assets, it is possible to re-use existing learning objects modelling, such as [5,21]. In the Pandora project, we are still in the process of selecting the most suitable modelling for our purpose.

There is a second type of stimuli for the simulation trainees. Since the sensing system records the “performance” of each participant also in terms of stress and anxiety, the simulation can be adapted to the specific conditions and deliver tailored inputs for the individual trainees. For example, if the purpose is to augment the pressure on a participant, the input could be made more dramatic. To this end, the Learning Delivery module makes use of *Non-Player Characters* (NPC): in games terminology, elements that act as a fictional agents and that are animated and controlled by the system. Those NPCs simulate additional actors from outside the learning environment and are used to deliver information to the trainees.

Our KB, therefore, includes also the modelling of NPC descriptions, in terms of their role in the simulation, their basic characteristics (e.g. gender, ethnicity, disability), their profiles (expertise, experience, emotional type, communication skills, etc.), their multimedia rendering mode (from the simplest text representation to fully rendered 3D avatar), etc. For this modelling, Linked Data are exploited for the reuse of pre-existing descriptions and Semantic Web technologies are leveraged to retrieve and select the most suitable NPC to simulate a desired stress or anxiety situation.

## 5 Exploiting Provenance Tracking

As detailed in the previous section, our Linked Knowledge Base is used to manage the knowledge required to produce simulation-based learning sessions. Simulation Learning can be seen as a special case of the Open Provenance Model (OPM) [18]. The sessions are our main *process*, the trainees, as well as the simulated external characters, are our *agents* and the events and the decisions taken by the trainees are the *artifacts* of the learning sessions.

Our investigation has recently focused on the definition of the suitable OPM Profile for Simulation Learning systems; specifically, we have mapped<sup>10</sup> our Timeline-based Planning Ontology to the Open Provenance Model Vocabulary Specification [25]. Hereafter we give some hints on how we can build on the Open Provenance Model and why it is useful.

The provenance tracking in simulation learning can be done at two levels: at design time – when the learning scenarios are modelled in the KB with their possible planning rules –, and after the learning sessions – when the results of the simulations are analysed.

At design time, provenance can be used to trace the cause-consequence chains between the simulation events. As explained in Section 4.2, planning rules are used to model the admissible transitions between events in the simulation; the *completion and inference rules* defined in OPM [18] can be exploited for the consistency checking of the simulation modelling. On the one hand, those rules can help in refining the modelling, by eliminating useless entities, combining eventual repetitions and introducing missing entities; on the other hand, OPM rules can help in examining the possible decision-trees (i.e., the possible alternative planning options) to identify unreachable states or decision bottlenecks.

We have identified a set of *causality checking rules*, expressible in terms of OPMV and of our planning ontology, and we have implemented them as a set of SPARQL 1.1 queries [13] on the simulation scenario models. Those rules represent a valuable support means for the scenario modellers: when the simulations are described in terms of the events and the planning rules to trigger the events, it can be difficult for the modeller to identify potential problems. It is worth noting that the causality checks we have identified are different and complementary to the planning consistency checks usually employed in planning systems [11].

After the learning sessions, the simulation records can be analysed to understand and synthetise the learning outcomes. Tracking the provenance of trainees' decisions and mining the most popular causal chains across several sessions delivery can be of great help for identifying learning needs, common behaviours (as well as common trainees' mistakes), wide-spread procedures, etc. This information can become of considerable importance: on the one hand, to improve the learning simulations and better address learners requirements and, on the other hand, to better study and interpret learning outcomes for individual participants or for entire classes of trainees.

---

<sup>10</sup> Cf. <http://swa.cefriel.it/ontologies/causality-provenance.html>

## 6 Conclusions

In this paper, we presented our approach and experience in building a Linked Knowledge Base to support Simulation Learning systems. We introduced the general architecture of such a system together with a concrete scenario in Crisis Management training; we illustrated the benefits of the use of Linked Data and Semantic Web technologies and we summarised our modelling choices. We also suggested the introduction of provenance tracking, to further enrich and better analyse the contents of a Knowledge Base for Simulation Learning.

Our approach is being integrated in the Pandora Environment, which, in the second half of 2011, will be tested at the UK Emergency Planning College in their “Emergency Response and Recovery” training courses.

**Acknowledgments.** This research is partially funded by the EU PANDORA project (FP7-ICT-2007-1-225387). We would like to thank the project partner for their collaboration.

## References

1. Aldrich, C.: Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to e-Learning, Pfeiffer (September 2003)
2. Berners-Lee, T.: Linked Data – W3C Design Issues, Architectural and philosophical points (2006), <http://www.w3.org/DesignIssues/LinkedData.html>
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems* 5, 1–22 (2009)
4. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia – A Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 7, 154–165 (2009)
5. Brase, J., Nejdil, W.: Ontologies and Metadata for eLearning. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies International Handbooks on Information Systems*, pp. 555–574. Springer, Heidelberg (2004)
6. Caird-Daley, A., Harris, D., Bessell, K., Lowe, M.: Training Decision Making using Serious Games. Tech. rep. Human Factors Integration Defence Technology Centre (2007)
7. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: *WWW 2005: Proceedings of the 14th International Conference on World Wide Web*, pp. 613–622. ACM (2005)
8. Celino, I., Dell’Aglío, D., De Benedictis, R., Grilli, S., Cesta, A.: Ontologies, rules and linked data to support Crisis Managers Training. *IEEE Learning Technology Newsletter, Special Issue Semantic Web Technologies for Technology Enhanced Learning* 13(1) (2011)
9. Cesta, A., Cortellessa, G., Fratini, S., Oddi, A.: Developing an End-to-End Planning Application from a Timeline Representation Framework. In: *21st Applications of Artificial Intelligence Conference* (2009)
10. Cesta, A., Fratini, S.: The Timeline Representation Framework as a Planning and Scheduling Software Development Environment. In: *27th Workshop of the UK Planning and Scheduling SIG* (2008)

11. Gerevini, A., Long, D.: Plan Constraints and Preferences in PDDL3. Tech. rep. R.T. 2005-08-47, Dipartimento di Elettronica per l'Automazione, Università degli Studi di Brescia (2005)
12. Gil, Y., Blythe, J.: Planet: A sharable and reusable ontology for representing plans. In: The AAAI - Workshop on Representational Issues for Real-World Planning Systems, pp. 28–33 (2000)
13. Harris, S., Seaborne, A.: SPARQL 1.1 Query Language. W3C Working Draft (2011), <http://www.w3.org/TR/sparql11-query/>
14. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. In: Synthesis Lectures on the Semantic Web: Theory and Technology, 1st edn., vol. 1. Morgan & Claypool (2011)
15. Hodgins, W., Duval, E.: Draft standard for learning technology - Learning Object Metadata. Tech. rep., Learning Technology Standards Committee of the IEEE. IEEE Standards Department, New York (July 2002)
16. Knublauch, H.: SPIN Modeling Vocabulary (October 20, 2009), <http://spinrdf.org/spin.html>
17. Lehto, M., Nah, F.: Decision-making Models and Decision Support. In: Handbook of Human Factors and Ergonomics, John Wiley & Sons, Inc., NY (2006)
18. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., den Busche, J.V.: The Open Provenance Model core specification (v1.1). Future Generation Computer Systems (2010)
19. Rajpathak, D., Motta, E.: An ontological formalization of the planning task. In: International Conference on Formal Ontology in Information Systems (FOIS 2004), pp. 305–316 (2004)
20. Snizek, J., Wilkins, D., Wadlington, P., Baumann, M.: Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions. *Journal of Management Information Systems* 18(4), 147–168 (2002)
21. Steinmetz, R., Seeberg, C.: Meta-information for Multimedia eLearning. In: Computer Science in Perspective, pp. 293–303 (2003)
22. Stern, E., Sundelius, B.: Crisis Management Europe: An Integrated Regional Research and Training Program. *International Studies Perspective* 3(1), 71–88 (2002)
23. Stojanovic, L., Staab, S., Studer, R.: eLearning based on the Semantic Web. In: WebNet 2001 - World Conference on the WWW and Internet, pp. 23–27 (2001)
24. Tiropanis, T., Davis, H.C., Millard, D.E., Weal, M.J.: Semantic Technologies for Learning and Teaching in the Web 2.0 Era. *IEEE Intelligent Systems* 24(6), 49–53 (2009)
25. Zhao, J.: Open Provenance Model Vocabulary Specification (October 2010), <http://purl.org/net/opmv/ns>