Linked Open Data Driven Game Generation

Rob Warren[1] and Erik Champion[2]

[1] Big Data Institute,
Dalhousie University,
Halifax, Canada
rhwarren@dal.ca
[2] School of Media, Culture and Creative Arts,
Curtin University,
WA, Australia
erik.champion@curtin.edu.au

Abstract. Linked Open Data provides a means of unified access to large and complex interconnected data sets that concern themselves with a surprising breath and depth of topics. This unified access in turn allows for the consumption of this data for modelling cultural heritage sites, historical events or creating serious games. In the following paper we present our work on simulating the terrain of a Great War battle using data from multiple Linked Open Data projects.

Keywords: Simulations, Consuming Linked Open Data, Serious Games.

1 Introduction

Linked Open Data (LOD, [1]) has created new avenues for content publishers to distribute their data while providing detailed linkages and annotations. We report here on a prototype method for automated Linked Open Data driven procedural game generation. This was a part of an interdisciplinary partnership between The Games Institute at the University of Waterloo, The Big Data Institute at Dalhousie University and the School of Media, Culture & Arts at Curtin University. The collaboration is the result of discussions on the generalizability and implementations of approaches such as Distant Reading [2], operational history and virtual heritage [3] using Big Data and Linked Open Data.

1.1 Simulation as an Information Retrieval Interface

The amount of information available digitally is increasing and the use of Linked Open Data approaches have made it a) available through standardized interfaces within structures that are self-documenting and b) these structures and their definitions can be linked across to different data sets.

These two items are most relevant in that they allow the retrieval of the information through a standardized interface while ensuring it relevancy through complex querying. Simulations, visualizations and games[1] have long been known

[1] In the course of this paper, we will use 3D simulation, visualization and game as interchangeable terms.

P. Mika et al. (Eds.) ISWC 2014, Part II, LNCS 8797, pp. 358–373, 2014.

to be effective means of communicating information. However their design and construction is a crafting process that is multidisciplinary, intuitive and does not lend itself easily to mass customization.

Because of the linkability and the built-in ontological support of linked open data, we believe the data-integration cost significantly lowered and that generalized information retrieval through simulation is now possible. As an example, the queries that unite Linked Geo Data, DBPedia and Muninn data to generate flora information in Section 4.2 can be hard-coded using conventional SQL and/or XML databases. However, the use of public facing SPARQL interfaces and OWL ontological constructs allow us to retrieve the data without requiring special access to the databases like conventional databases would.

Google Earth is a analogous example of a standardized visualization engine that represents data from multiple concurrent sources each with its own API, data definitions and administrative process. The ongoing data reach of Google Earth is only possible through the large ecosystem supported by Google that ensures the ongoing maintenance of the data-source specific translation mechanisms.

The objective of this research direction is not to replace the simulation designer, but rather automate a number of processes that are fundamentally sophisticated information retrieval and processing. This automation has already occurred with areas such as game physics which are now mostly handled by the game engine instead of the game designer. This is the next logical step as we move from one-off game designs to game engines, to manual content generation, to the current procedural game generation and then to this proposed linked open data-driven game generation.

This paper reviews our initial attempt at creating a generalizable engine capable of making use of Linked Open Data to construct a simulation. We also state that these simulations as realistic not because of the visual accuracy of the rendering, but in that the events, places and things that occur within the simulation are based on documented facts.

The organisation of the paper is as follows: we first discuss previous work and the benefits of using LOD for simulations, followed by a description of a prototype based on the events of the Great War using data from the Muninn Project, a LOD project focused on the Great War. We then review some of the lessons learned from the exercise and we close with a discussion of ongoing work on the methodologies needed to consume data in context.

2 Previous Work

Data driven simulations are not new and in common use with Building Information Management [4], Cultural Heritage [5] and even recreational game development. Concurrently, Linked Open Data has been used extensively for describing archival [6] and bibliographic [7] material and as well as GIS data [8]. A renewed interest in the ideas of the Venice Charter is driving the markup and storage of Cultural Heritage data using LOD formats owing to its ability to record data in a long lasting description format.

Linked Open Data is seen as a desirable technology for Digital Humanities especially those focused on the Galleries, Libraries, Archives and Museums (GLAM) industries. LOD for cultural heritage allows updateable information, closer collaboration with archival institutes and more responsive design for different platforms while being aligned with some of ideas of the London Charter on cultural heritage visualization.

Using LOD to procedurally create content directly from a live database on a as-needed basis is the next logical step that builds on previous works on procedural content generation and the Semantic Virtual Environment [9]. From the game generation perspective, a taxonomy of possible methods is reviewed by Togelius et al. [10] with some early work on [11,12] Procedural Content Generation (PCG). Kallman and Thalman [13] proposed the precursor of current "prefabricated" objects within games by having an event model where different objects could interact with one another. Müller et al. [14] and Andrés et al. [15] also using combined photogrammetric and procedural modelling to build models of a city in a automatable fashion.

Tutenel et al. [16] primarily saw the semantics problem of the data as a means of tracking the contents within games and keeping track of the physical constraints at design time. Vanacken et al. [17] used a similar approach using Ontology Web Language (OWL) to track objects within the virtual world and their properties. The Semantic Virtual Environment (SVE) was proposed by Otto [9] where the minutia of the virtual environment was linked to a semantic web database holding the properties and hierarchy of the environment. The intent was to have a higher level knowledge system of the items within the virtual environment using inheritance of properties between object classes. Fuhrmann et al. [18] designed an ontology to record the appearance and organisation of clothing within a virtual environment so that it could be procedurally generated while Gutiérrez et al. [19] human beings ontology where a person's motions were related to ensure realistic motion. Games With A Purpose (GWAP) have also been proposed [20] as a means of acquiring ontological or annotation data. Most recently Fribeger and Togelius [21] made use of LOD to create a Monopoly-like board game customized to a locality. Reitmayr et al. [22] used a similar concept with augmented reality to relate objects within the world to external data sources, such as web pages. Grimaldo et al. [23] saw the semantics and ontological aspects primarily as a means of negotiating relationships and behaviours between objects within the virtual world.

A large part of procedural game design was the simulation and creation of "game level" as outlined by Tutenel et al. in 2008 [24]. While the procedural generation of game terrain through the detailed simulation of an ecosystem, as described by Dussel et al. [25] can be simulated from scratch, the trade-off between random simulation and data-driven parameter simulation is based on the data available for the simulation. In a similar vein, Lu et al. [26] reviewed simulation of weathering, rusting and cracking of paint based on both computation simulations of the break of the material and the emulation of the visual aspect of the cracking.

Trinh et al. [27] wrote about the use of the semantic web for the representation of orientation, direction and attitude within a a virtual world. Coyne et al. [28] had a similar approach that was used to translate contextual descriptions of a world into a virtual materialization based on the use of constraints to disambiguate the world objects and their positions.

To the best of our knowledge the use of Linked Open Data to automated the creation of data-driven procedurally generated 3D simulations of historical and current locations is a novel contribution. Previous contributions were primarily tools supporting the designer in creating an environment that would then be statically used within a game. A secondary contribution is the creation of culturally sensitive prefabricated assets within a simulation environment that taken on the content appropriate to the both location and time.

3 Theoretical Framework

A significant amount of data exists encoded in Resource Description Framework (RDF) and Ontology Web Language (OWL) formats on the web about topics ranging from geography to linguistics. Unlike previous approaches to data markup and storage, the detail contained within LOD has reached sufficient complexity that it is now difficult for a human to enumerate, let alone query, LOD.

Creating a generalizable data-driven simulation implies the following two types of queries:

a) Well defined, structured queries that return deterministic results based on known parameters. These include spatial and temporal information to which a number of other events and objects are tied to, such as the sun, the moon, terrain, etc. Some uncertainty may be contained within this process but the actual query process itself remains deterministic.

b) Poorly defined, serendipitous queries whose goals are ancillary to the era or the story telling aspect of the simulation and that create the minor details that are relevant to its realism. This can include "litter" on the ground, posters, titles of books on shelves and so on.

The use of ontologies and databases is ideal for this purpose in that the ontological description of a thing enables the simulation engine to determine what properties should be located to materialize an instance. A more generalizable approach is the use of a simulation; the detailed minutia of information lends itself well to automated querying of the information necessary to answering a hypothesis.

A secondary benefit is that the ontology can also be used to determine what details can be omitted from a materialized instance and which can be estimated. It also gives guidance to the simulation engine as to how to query the database in order to estimate a missing property. As an example, a military trench on a battlefield has been dug into the ground at a depth and width. In many cases geometry information or archival information will reveal the dimension of a specific trench and it can be rendered. When the dimensions of the trench are

not known, they can be replaced by the statistical average width and depth of trenches within the area.

4 Simulating the Great War

Given the centenary of the Great War and the primary author's ongoing work in this era, an area of Western France in late 1914 was chosen as case study to simulate[2]. Later on in 1917, this would be the site of the Battle of Vimy Ridge, which would see the Canadian Corps fight as a single entity for the first time. This location is ideal as a test case for a number of reasons including the presence of two distinct areas (Entente trenches versus Central trenches) with different cultures and features changing over time.

Fig. 1. A modern topographical map of Vimy Ridge (OpenTopomap)

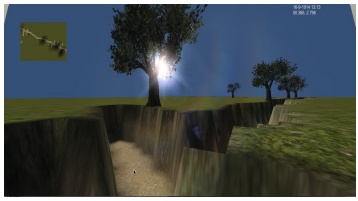

Fig. 2. Vimy Ridge on September 16, 1914 at about 50.366N, 2.796E

An example use case for this methodology would be an historian attempting to determine who had the advantage of the terrain shown in Figure 1. Given this map, an experienced historian or geographer could make an educated evaluation based on his interpretation of a two dimensional document. If the contents of the map are digitized into a GIS data structure and merged with other information, a programmer could design a one-off algorithm to try and determine terrain advantage. No doubt, this would require extensive consultation with the historian and a significant amount of time wrestling with deep methodological questions.

Within a simulated environment based on the actual digital terrain elevation and trench data, a historian can simply walk around in the virtual world and have direct understanding of the advantages of terrain. The argumentation and narrative proof of the conclusion conducted by the researcher has not changed from previous methodologies but the ease of access and interpretation of a highly described information source is dramatically improved. Linked Open Data also allows us to promote the convergence of multiple consumption and analysis strategies.

We review here four aspects of the simulation generator based on the issues that they identify: the layout of the trenches on the battlefield, the elements of the vegetation estimation, the use of culturally attuned prefabricated objects and the crediting of work and data sources within the simulation. The simulation was

[2] `http://rdf.muninn-project.org/demo`

implemented with the Unity game engine drawing data from LOD databases. The approach is meant to use multiple concurrent SPARQL servers, however given Unity's lack of Cross Origin Request Security (CORS) support, we use a intermediary SPARQL server that proxies the retrieval of data from multiple LOD servers.

4.1 Trenches

The First World War in Western Europe was a war of attrition fought in trenches that were often within earshot of the enemy. In this section we review the creation of the terrain through geometries that are obtained through the Muninn project. The geometries and features are derived from British Trench Maps in a Linked Geo[8] format that also support access using the Ordered List Ontology, further details have been previously published in [29]. Figure 2 is a screenshot of the simulated trenches. The information used to generate the trenches includes both depth and width, however additional information about the state of the trench (abandoned, occupied) is also available and effects how the trench is represented.

The state of an object, or its serviceability can be recorded in one of two ways: a state property or the use of a class hierarchy as Linked Geo Data does. An example is RailwayThing which is the superclass for Rail, LightRail and AbandonedRailway. Muninn makes use of the graves:hasState property that allows us to record the state of the feature as reported on the original British Trench Maps. The reason for this decision was that while the state of a real world object changes over time, its fundamental identity does not and the use of sub-classes to encode state results in an unmanageable number of classes.

This state is what the simulation engine uses to make decisions on how the LOD is translated into the visuals, materials and terrain. In the case of the trenches, 'abandoned' trenches will be filled with debris and the width and depth of the trenches will be reduced by a small, randomly changing percentage to simulate neglect.

This process of simulating a trench is executed by depressing the terrain in the median line of the trench as reported by the LOD. Width is set according to each position node within the LOD stream and the textures of the trench bottom, walls and fields set according to generic Unity textures.

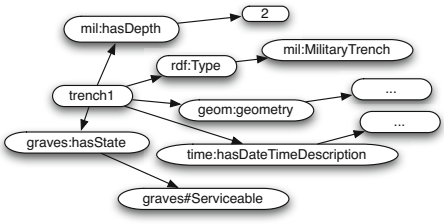

Fig. 3. Basic ontological view of trench data

4.2 Vegetation

We review here the generation of vegetation as recorded in archival and GIS information sources. Different data sources report information on vegetation at different levels of granularity over time. Trees and forested areas are created based on the basis of available and the processes described in this section are applicable to shrubs, bushes, hedges and plantation fields.

When inserting new vegetation in the simulation, select 3D assets or textures that fit the particular class of vegetation is not difficult given the availability of 3D assets. The difficulty lies in managing under-specified vegetation instances or aggregates of vegetation instances such as a forest. In some cases, an area may be a tree plantation of a very specific species, in others an old-growth forest and in some cases, individual trees.

What species of tree, tree size, height and any seasonal effects on appearance will dictate appearance and modification of generic assets. Furthermore, in the case of historical data the vegetation information may only be available for certain points in time and any visualisation in between these points has to estimate the state of the information by interpolation or other means.

Within our prototype simulation, trees are programmatically generated based on reported flora information from British Trench Maps. These maps provide an interesting challenge as in this era they were printed using a generic base plate from one date while coloured overlays were added with the most up-to-date information. Both layers represent information at different points in time where not all features are up to date.

The state of the features is an interesting problem: it is obvious that trees within a battlefield would be heavily scarred by projectiles and shrubbery would have been destroyed by the movement of personnel and pack animals. Additional statistical simulations driven from external event data, such as battles, could be used to infer the amount of damage that would occur. In this case we chose not to implement such a non-trivial process but note that there may be some opportunities for ontologically inferring the state of a feature based on the presence of a second feature: that a ground track has been indicated on a map implies that it can be visually differentiated from the ground next to it and thus the state of the general area can also be determined.

As with the previous section on Trenches, an ontological approach to data management has some advantages under uncertainty. The ontology can be reused to locate or estimate properties that are missing from the source data or derived from other ontological constructs. Consider a simple tree ontology T that presents the following properties for each tree instance: h height from the ground, c height of the crown of the tree and the width w of the crown[3]. If a single tree does not have size or species information specified, it can be estimated by querying neighbouring trees. A simple query of neighbouring $G_{Avg(h)}$ $\rho_{Dist(Lat,Lot)<Radius}$ $(t \in T)$ trees can estimate height h or use a frequency

[3] These measurements were chosen because of their simplicity and their direct use in classifying trees using LIDAR data [30].

approach for a species s for a value or $\pi_{s_i}\ G_{Max(Count(s))}\ \rho_{Dist(Lat,Lot)<Radius}$ $(t \in T)$ to select a species instance s_i of the species class S.

Another means of estimating foliage data is to infer its properties based on other data-sets. Given the coordinates of the tree and the month of the year, one should be able to find LOD about foliage properties and the regions in which they thrive. DBpedia presented a ready made collection of information against which queries could be run to answer the basic question: *at a certain time and place, if a tree were to exists, what would it be and how would it look like?*. The triple pattern in Figure 4 represents the graph required to locate the tree species probable in the area being simulated.

```
dbpedia:Flora_by_country  skos:broader      ?f .
?f                        dcterms:subject  ?tree .
Category:Ornamental_trees dcterms:subject  ?tree .
Category:Trees            skos:broader     Category:Ornamental_trees .
```

Fig. 4. Finding the right species of ?tree based on a regional category ?f

Binding ?f should be Category:Flora_of_France for the right set of ?tree's to be retrieved. However there exists no term that links country or regional terms to these categories, limiting the generalization of the approach: in an ideal case, we should be able to translate coordinates to a country or region to a tree category.

A similar experience occurred when querying ontologies and data-sets dedicated to flora and forestry. Designed and authored by domain specialists, these data-sets are deeply integrated into the biology application ecosystems and descend directly from human readable, portal published databases. Tightly coupled to their primary objectives and have no connections to generalizable concepts that would frame their data to solve other problems.

For example, the Plant Ontology data-set has for mandate to be a "controlled, structured vocabulary (ontology) of terms to describe plant anatomy, morphology and the stages of plant development". Its contents are effectively a taxonomy for biological classification using Binomial nomenclature. As per accepted ontology best practices, it lacks a generalizable object that would relate its terms to dbpedia:Tree or dbpedia:Plant as this would be a "lazy concept" [31]. While appropriate in the pure theoretical ontological view, these design decisions make ontological discovery and matching impossible. Thus, in our experimental implementation we chose to use dbpedia as a data source for our fauna information.

Figure 5 is a screenshot of what occurs when a user clicks on a specific tree in the simulation. The information about the tree is reported as well as the provenance of its information and the decision mechanism that was used for its placement, shape and species.

4.3 Culturally Aware Prefabricated Game Assets

Realistic simulations can contain 3D assets that represent everyday objects. These are not directly relevant to the simulation or information need but help

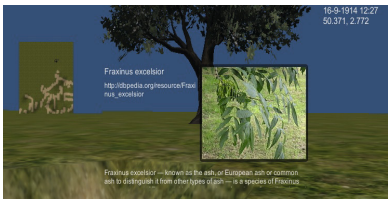

Fig. 5. Objects such as individual trees are created based on actual data. The simulation provides both the individual tree instance and tree class information.

Fig. 6. A culturally aware phonograph in the Canadian trenches. It's placement is purely demonstrative, but the content played is relevant to the locality.

create a comfortable aesthetic. Within the framework of Section 3, these queries are poorly defined as their objective is to be sensical without necessarily being relevant to the initial query.

Designers usually make these assets to be simple image placeholders that carry no information beyond appearing realistic. With a LOD approach, the creation of these assets can be mechanized through the retrieval of contents from Linked Open databases relevant to the time or place. We note that projects, such as CARARE [32] are working on creating repositories of 3D objects that could be expanded with these specific behaviours.

In our simulation, we chose to focus on documents and media that are described in period texts as littering the trenches of a battlefields such as discarded newspapers or postcards. Dugouts are similarly documented as containing creature comforts such as books, posters, paperwork or maps.

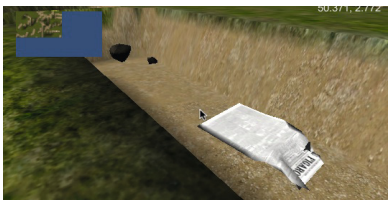

Fig. 7. A discarded *Figaro* newspaper in the *Entente* trenches.

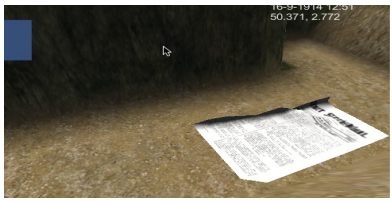

Fig. 8. A discarded *Gumbinner Kreisblatt* newspaper in the *Central* trenches.

Figures 7 and 8 are screen shots of discarded newspapers that move in the wind across a battlefield. The newspaper is a prefabricated asset ("prefab") that can be instantiated anywhere on the battlefield but which displays a random page from a newspaper that is appropriate to the location or nationalities nearby on a date that is weeks beforehand the simulations. This is made possible by partial LOD made available by the French and German National Libraries of their holdings.

The same can be done with books whose contents and covering images are available from LOD data-sets such as Project Gutenberg or Archive.org. Another interesting prefabricated asset that was developed for this project was that of a phonograph, an item which would be commonly available in some of the rear areas of the battlefield to entertain the troops. The phonograph 3D asset of Figure 6 is one that is freely available, through behaviour has been added which is similar to that of the newspaper 3D asset: songs and airs related to nearby locations or nationalities are chosen from the appropriate era. As with the example of Figure 5, clicking on the phonograph reveals the LOD resources used by the 3D asset. These types of 3D assets supported by LOD data sets are desirable because they add realism relevant to the era without requiring further localisation work.

4.4 Credit Generation

Lastly, the use of Linked Open Data entails that some measure of provenance is available to tell us about every single term retrieved from servers. This opens up the interesting possibility of creating a bibliography, citation list or credit listing for both the creators and sources that were involved (in)directly in the simulation. The traditional notion of authorship is now challenged as the increase in the complexity of creative works [33] makes assigning credit and authorship non-trivial.

In this case, our simulation creates a new LOD document that links to all of the URL's retrieved to create the simulation. This document is then parsed when the simulation is exited and used to create a rolling credits scene that outlines all data sources used by the simulation. Individual roles played in its construction can be recovered from a series of statements attached to the consumed LOD. The question then becomes how much detail should be shown in the credits: should every contributor to a Wikipedia page be credited for the use of a DBPedia term? Should the order of the credits be based on the relative importance of each contribution or on its chronological use within the simulation?

The current generation of credit sequence is driven by the number of terms used from each source, with the intuition being that the data-set used most often is the most important. As an added aesthetic touch, the media assets (images and music) can be randomly played in the background while credits move up. This serves to remind the audience that the individual assets used by the simulation were retrieved from several sources and were not statically chosen.

5 Experimental Results

In this section the performance of our experimental implementation is reviewed and conclusions drawn from our use of Linked Open Data.

5.1 Performance Evaluation

One of the concerns with online procedural content generation was the speed with which scenes could be generated from Linked Open Data. This is a valid

concern given the amount of information needed to generate an ad-hoc scene and the use of the SPARQL query language which is often criticised as being unresponsive.

In practice, the retrieval time of information from different LOD databases is not a concern, even with consumer grade internet connections. A disproportionate amount of wall-clock time is spent in creating the virtual world within the simulation engine. The amount of detail that linked open data driven procedural generation entails is high and for a few LOD entities, several dozen operations on a specific area of the virtual world need to be performed and prefabricated objects instantiated.

Procedural content generation has traditionally been a design tool used to create a terrain or world that is then statically stored for use. In this case, the LOD is retrieved directly from the endpoint and a new terrain is generated from the most up-to-date data available at run time.

The use of SPARQL servers over that of APIs was a necessity in that they provide standardized access to information using a generalizable query language. In some cases, the retrieval of content could only be done through other APIs or by parsing the HTML code of web portals. This situation should be remediated as it limits the ability of the content consumers to search multiple information sources without significant investment for each additional source. The Museum API[4] Page lists over 50 different APIs of historical interest. The diversity of methods, formats and specifications becomes a hindrance to a generalizable method as each new data source must be integrated and maintained individually.

Accuracy is a thorny problem which remains an open area of research. Domain expert and layman interpretation of the same visualization can be different as well as their requirements. For example, in our demonstration simulation we randomly place discarded newspapers as described in Section 4.3. While first hand accounts of the war tell us that discarded newspapers were common and the date and instance of the newspaper respects locality, its placement is random and only for aesthetic purposes. Depending on the audience, one can argue that this is needed for communicative purposes or deliberately misleading to the audience.

In our prototype, every prefabricated object placed within the environment can be selected and a pop-up window will report on its provenance, instance and class, as in Figure 5. This type of reporting is an effective means of providing justification or explanation on the asset placement and has been previously been used by Pauwels et al. [34] in documenting engineering drawings.

This works well for discrete objects such as basic shapes like individual trees but poorly for complex structures such as buildings or terrain surfaces where the object of interest is hard to select. Interestingly, from the perspective of creating a simulation, there is no difference between a historical simulation and a current-world simulation as the uncertainty that is represented within the LOD is dealt with in the same way. While not done here for lack of space, we can change the location and time of the simulation for a current day metropolitan city and display a simulation without problem.

[4] `http://museum-api.pbworks.com/Museum%C2%A0APIs`

5.2 Insufficient Provenance Information

An element in historical reconstruction and building information management is the tracking of the provenance of information both in terms of information provider and creation process. Currently, semantic web terms can be related to their home URI through the rdfs:isDefinedBy, a data-set description void:inDataset, a "source" dc:source or to detailed provenance information with the PROV ontology.

In attempting to locate data sources for this project, few data-sets were found that had detailed provenance, as confirmed independently by Buil-Aranda et al [35]. The underlying assumption is that the data-set is already known and explicitly queried by a particular user or his agent. There is anecdotal evidence that suggests the reason for this is that the ontology or data-set is created with a specific community in mind and no outside use was foreseen.

This lack of overall description is a concern in that this makes the automated discovery of additional data in ways similar to that of Akar et al. [36] impossible. This prevents tracking the authorship or ownership of data that would make citation straightforwards and credit sharing possible. But foremost, it prevents the sharing of quality, precision and process information that should apply to all terms of the data-set, such as the date of issue or validity. This information is especially necessary when dealing historical data as the data set description may be the only means of assigning temporal information to the remaining terms.

As an example, DBPedia has attempted to document its processes and the raw source used to created a dbpedia term through the use of the dc:source term linked to the original Wikipedia page. It is not possible to gleam publication date for the resource without parsing the HTML source or already knowing the location of the data set description since no void:inDataset is provided to link back to it.

When using LOD that moves beyond small, known data sets the necessity of including rdfs:isDefinedBy and void:inDataset terms with every resource becomes evident. The reason for this is that the necessary use of aggregating and triple search services obfuscates the original provenance of the information, including by changing the base of the URL. We did notice a bias in data-set authorship in that having one's data linked to (cited) is valuable, while linking to (citing) another data-set less so. Some of the most linked to data-sets, are also the least likely to link to other data-sets. Part of this is due to the concern that their data might be tainted with inaccuracies. Of course, linking (citing) requires effort on the part of the data-set author while the effort of being linked to is externalized to others. This classic agency theory problem remains to be solved within LOD systems.

5.3 Semantic Versus Appearance

The semantic web is meant to provide semantic information to a thing as opposed to the World Wide Web which provides appearance information to a thing. It therefore comes as a concerns that a number of practices meant to support human

consumption of the data are hindering machine consumption of the data. An example of this is the use of basic descriptive terms that contain content which are mapped directly from text fields from a catalogue entry. This is due to data providers using a "mapping approach" to creating LOD from existing databases. This is a problem in that the fields in these initial systems were meant to be human readable, but are of little use to a machine.

This has primarily to do with the fact that most archival LOD is being generated through a "mapping" approach to creating LOD that does not necessarily translate classifications as linkages and simply copies strings. In many cases a search for a certain type of publication, such as a newspaper, is not possible unless the correct human-readable string is used. Some cataloguing implementations use the the original description strings to record publication types, such as <dc:type>czasopismo</dc:type>, requiring a prior knowledge of the national language before the data can be queried. This shows a deep-seated assumption that content is still retrieved by keywords or that the RDF properties contents are only meant as displayable content.

The above case makes locating newspapers impossible unless searching for every translation of the string "newspaper" in use within the data-set. The concern is that the considerable efforts that have been made in using standardized cataloguing processes are being wasted by simply "mapping" the string values into LOD terms without appropriate pre-processing. Another example of the problems created by direct mapping occurs when using audio recording data from the PCDHN data-set. Here recordings are represented by LOD structures that represent a concept, its manifestations and the actual recordings at items. However, it is impossible to programmatically determine what language the songs are actually interpreted in, and whether the materialization of the concept is the whole record or a single track. This materialized expression has the property<dc:subject xml:lang="fr">Chansons française</dc:subject>, which is ambiguous as to whether the song is culturally French or performed in French. Lastly, the two items linked to the manifestation are in two different binary formats which require either parsing the extension type or connecting to the web server until an acceptable encoding is found, requiring extra exemption handling on the part of the client. The problematic structure of the LOD is due to an archival view of the data where the approach of LinkedBrains / MusicBrains to representing audio recordings may be preferable.

The same concerns about mapping human readable data apply to Library of Congress (LOC) headings and terms, where *chansons françaises* makes no differentiation between the cultural origins, the localization or the specific interpretation of the works. In trying to locate content that is localised both linguistically and culturally these are important concerns. The problem is compounded by the use of strings, which need to be both culturally localised and transliterated.

The aggressive use of SKOS vocabularies, without supporting OWL statements, makes the discovery of additional resources and their query difficult in that the taxonomy must be known before the SPARQL query is written. Worse, authoritative vocabularies such as the LOC Subject Headings have traditionally

had multiple top-level concepts that have intermingling hierarchies. This means that concept Tree has skos:broader concepts Nursery stock, Woody plants and skos:narrower concepts Bark peeling and Fruit trees. Primarily meant for human consumption, subject headings were never meant as a hierarchy of concepts and the implicit change in conceptSchemes prevents it from being used for automated querying, even in a non-transitive context. This same model of "multiple trees" is used by DBPedia to translate Wikipedia category data into LOD and highlights some of the hurdles in transforming human generated data to machine readable data.

5.4 Separation of Meta-data and Data

One of the early users of LOD has been the GLAM communities that see a flexible means of distributing data to their clients and amongst themselves. One of the challenges of consuming this data is the underlying assumption that LOD is meant as a meta-data framework, separate from the document.

This reflects the strong cataloguing tradition of these institution in a pre-digitization environment where interacting with the document (eg: a book) was a separate act from its discovery in a catalogue. In some cases online LOD sources have the document available in digital format but it is segregated in a separate system that is neither machine readable nor linked to by the LOD catalogue.

It would be preferable if both data and meta-data were represented in the LOD set as this would enable not only seamless access but opportunities for supporting data mining projects. Furthermore we note that HTTP content negotiation as it is currently used to support different LOD representations (See [37]) can also be applied to media formats. As an example, the Project Gutenberg RDF catalogue makes extensive use of dc:hasFormat term to link a document to the different file formats it is available under. It is desirable, especially for human consumption, to have different URL's that allow a client to explicitly specify a format. But given the sophistication that is expected from a LOD client, and indeed that is already present in the average Web browser, offloading media type negotiation to the HTTP transport layer seems reasonable.

The Muninn Project currently makes use of content negotiation to not only select the appropriate LOD serialization but also serve images in the format requested or accepted by the browser. Similarly, the Stanford IIIF JSON-LD API also uses the same approach in its image content negotiations. The unification and simplification in the number of APIs and endpoints needed to reach the content needed is desirable from both an efficiency point of vue and to lower the cost of development.

6 Conclusions

In this paper we presented a novel, generalizable way of consuming LOD within a game engine to create on-the-fly digital simulations of historical or present day places. We also presented the novel contribution of culturally adaptable 3D

prefabricated assets that can take on the content or appearance appropriate to time and location.

The following issues were noted in consuming LOD for 3D virtual worlds: a) LOD is being published under the assumption that it will be consumed by humans, sometimes making machine consumption impossible, b) data-set authors are neglecting to provide provenance and data-set information that would allow proper data-set discovery and c) some LOD publishers still see publication as a meta-data search mechanism for other services which requires unnecessary retrieval steps on the client's part.

In future work, we will focus on scenario and story generation using detailed event data found within Linked Open Data databases as well as generalizing the extraction of behaviours from generic objects within the virtual world.

References

1. 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)
2. Moretti, F.: Distant Reading. Verso (2013)
3. Gaitatzes, A., et al.: Reviving the past: Cultural heritage meets virtual reality. In: Conf. Virtual Reality, Archeology, and Cultural Heritage, pp. 103–110 (2001)
4. Fai, S., Graham, K., Duckworth, T., et al.: Building information modelling and heritage documentation. In: CIPA International Symposium, vol. 47 (2011)
5. Anderson, E., McLoughlin, L., et al.: Developing serious games for cultural heritage: a state-of-the-art review. Virtual Reality 14, 255–275 (2010)
6. Mazzini, S., Ricci, F.: EAC-CPF ontology and linked archival data. In: Workshop on Semantic Digital Archives, pp. 72–81 (2011)
7. Ford, K.: Lc's bibliographic framework initiative and the attractiveness of linked data. Information Standards Quarterly 24, 46–50 (2012)
8. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: Linkedgeodata: A core for a web of spatial open data. Semantic Web Journal 3, 333–354 (2012)
9. Otto, K.A.: The semantics of multi-user virtual environments. In: Workshop Towards Semantic Virtual Environments, pp. 35–39 (2005)
10. Togelius, J., Yannakakis, G.N., Stanley, K.O., Browne, C.: Search-based procedural content generation. In: Di Chio, C., et al. (eds.) EvoApplicatons 2010, Part I. LNCS, vol. 6024, pp. 141–150. Springer, Heidelberg (2010)
11. Togelius, J., Preuss, M., et al.: Towards multiobjective procedural map generation. In: Workshop on Procedural Content Generation in Games, p. 3 (2010)
12. Hartsook, K., et al.: Toward supporting stories with procedurally generated game worlds. In: Computational Intelligence and Games, pp. 297–304 (2011)
13. Kallmann, M., Thalmann, D.: Modeling objects for interaction tasks. In: Proc. Eurographics Workshop on Animation and Simulation, pp. 73–86 (1998)
14. Arisona, S., et al.: Increasing detail of 3d models through combined photogrammetric and procedural modelling. GIS 16, 45–53 (2013)
15. Andrés, A.N., Pozuelo, F.B., et al.: Generation of virtual models of cultural heritage. Journal of Cultural Heritage 13, 103–106 (2012)
16. Tutenel, T., Smelik, R.M., et al.: Using semantics to improve the design of game worlds. In: Artificial Intelligence for Interactive Digital Entertainment (2009)
17. Vanacken, L., et al.: Semantic information during conceptual modelling of interaction for virt. env. In: Multimodal Intf. in Semantic Interaction, pp. 17–24 (2007)

18. Fuhrmann, A., Groß, C., Weber, A.: Ontologies for virtual garments. In: Workshop towards Semantic Virtual Environments, pp. 101–109 (2005)
19. Gutiérrez, M., García-Rojas, A., et al.: An ontology of virtual humans: Incorporating semantics into human shapes. Visual Computer 23, 207–218 (2007)
20. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 319–326 (2004)
21. Friberger, M.G., Togelius, J.: Generating game content from open data. In: Conference on the Foundations of Digital Games, pp. 290–291 (2012)
22. Reitmayr, G., Schmalstieg, D.: Semantic world models for ubiquitous augmented reality. In: Workshop towards Semantic Virtual Environments (2005)
23. Grimaldo, F., Barber, F., et al.: Semantic virtual environments for interactive planning agents. In: International Digital Games Conference, vol. 17 (2006)
24. Tutenel, T., Bidarra, R., Smelik, R.M., et al.: The role of semantics in games and simulations. Computers in Entertainment 6, 57 (2008)
25. Deussen, O., et al.: Realistic modeling and rendering of plant ecosystems. In: Conference on Computer Graphics and Interactive Techniques, pp. 275–286 (1998)
26. Lu, J., Georghiades, A.S., Glaser, A., Wu, H., Wei, L.Y., Guo, B., Dorsey, J., Rushmeier, H.: Context-aware textures. ACM Trans. Graph. 26 (2007)
27. Trinh, T.H., et al.: Integrating semantic directional relationships into virt. environments. In: Virt. Env. & 3rd Joint Virt. Reality, pp. 67–74 (2011)
28. Coyne, B., Sproat, R.: Wordseye: an automatic text-to-scene conversion system. In: Computer Graphics and Interactive Techniques, pp. 487–496 (2001)
29. Warren, R.H., Evans, D.: From the trenches - API issues in linked geo data. In: Linking Geospatial Data Workshop, W3C (2014)
30. Ko, C., Sohn, G., et al.: Tree genera classification with geometric features from high-density airborne lidar. Cdn. Jour. of Remote Sensing 39, S73–S85 (2013)
31. Gherasim, T., Berio, G., Harzallah, M., Kuntz, P., et al.: Problems impacting the quality of automatically built ontologies. In: Proceedings of KESE, vol. 949 (2012)
32. D'Andrea, A., Niccolucci, F., Fernie, K.: Carare 2.0: a metadata schema for 3d cultural objects. In: Digital Heritage 2013. IEEE (2013)
33. Horton, R., Smith, R.: Time to redefine authorship: A conference to do so. BMJ: British Medical Journal 312, 723 (1996)
34. Pauwels, P., et al.: Linking a game engine environment to architectural information on the semantic web. Journal of Civil Eng. and Arch. 5, 787–798 (2011)
35. Buil-Aranda, C., Hogan, A., Umbrich, J., Vandenbussche, P.-Y.: Sparql web-querying infrastructure: Ready for action? In: Alani, H., et al. (eds.) ISWC 2013, Part II. LNCS, vol. 8219, pp. 277–293. Springer, Heidelberg (2013)
36. Akar, Z., Halaç, T.G., Ekinci, E.E., Dikenelli, O.: Querying the web of interlinked datasets using void descriptions. In: Linked Data on the Web (2012)
37. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synthesis Lectures on the Semantic Web: Theory and Technology 1, 1–126 (2011)