

Ontological specification of an authoring interface for creating sustainable language learning content.

Jozef COLPAERT

IOIW, Universiteit Antwerpen, Belgium

jozef.colpaert@ua.ac.be

Abstract: The creation of language learning materials is very labor-intensive, but a lot of learning content gets lost at every change due to its inability to adapt to new technologies, products, services, pedagogical models and educational demands. For the educational publishing sector, this low reusability or lack of sustainability, combined with the high authoring cost, certainly for a small market like Flanders, seriously hampers the profitability, if not already the viability, of the activity. Learning content, on the other hand, is vital for the quality of education and, indirectly, for society in general.

The objective of our current research is to deliver an ontological specification of an authoring interface for creating sustainable language learning content based on the identification of psychological requirements on the one hand, and on solving technological issues on the other. It is based on our object model for software architecture and database structuring: a model which is the result of more than 20 years of research and development in the field, and which has been thoroughly validated theoretically and empirically over the last couple of years.

The result should be an ontological specification for publishers: the entire project should yield a significant reduction of production cost on the one hand, and an increase in sales on the other by opening up both their product range and their markets.

Keywords: Sustainable learning content, generic database structures, authoring of language learning content, user interfacing

1. Introduction

In this article we will discuss the requirements which should be taken into account when designing an authoring interface for developing sustainable language learning content. Based on an already developed and implemented generic object model, currently being tested on technological requirements, our endeavor is to deliver an ontological specification which can then lead to subsequent research and development projects.

The impact of education on society, on cultural life, scientific achievements and economic prosperity can never be overestimated. Trying to improve the quality of education is or should be a lifetime challenge for educational researchers, teachers and policy makers. Educational research focuses on aspects such as learning styles, learner motivation, evaluation, class management or school policy. The Institute for Education and Information Sciences (IOIW) at the University of Antwerp is, since 2008, slowly gaining a solid reputation in this area. Not only theoretical aspects, but also more applied issues in the fields of educational technology, instructional design and engineering deserve our attention. We try to focus on topics where there is an obvious need on the one hand (a clear demand from educational reality) and a lack of available knowledge (a clear need for epistemological contribution) on the other.

The factors which triggered our attention were a/ a clear problem related to *learning content* which will be described in more detail below, b/ the fact that there is a clear lack of literature and findings on the topic and c/ our vision based on our track record since 1986 in the field of educational

technology and the production of learning materials (Colpaert & Decoo, 1999). This research should be seen as a stage in the empirical and theoretical validation of this vision as working hypothesis in an engineering approach.

By using the term learning content, we do not only refer to traditional textbooks, but also to materials such as syllabi and handouts, interactive exercises in applications like *Hot Potatoes* or *QuestionMark*, course content in electronic learning environments like *Blackboard*, video and sound clips (podcasts), presentation slides in *Powerpoint* or *Prezi*, materials for Interactive Whiteboards, web pages, wikis and e-reader content. The creation of this learning content is a very specific, labor-intensive activity for the following reasons:

- Writing learning content does not happen in a linear way, but it involves an arduous cyclic process of creating, editing, combining, structuring and formatting materials in several layers.
- Content should comply with many pedagogical, linguistic and cultural requirements such as to be linguistically correct, adapted to a specific level and context, engaging and attractive, as interactive and relevant as possible and to be politically correct by avoiding any statements or images which could insult or irritate individuals or minorities.
- Most authors work with a text processor like *Word* or with a Desktop Publishing program like *QuarkXPress*, *Adobe Indesign* or *Microsoft Publisher*, but to our knowledge there are no tools developed by design for assisting developers of learning content, teachers and/or authors in creating learning content and for speeding up the production process.
- It is not easy to retrieve, select, evaluate and integrate materials developed by others, due to the fact that they are protected by copyright, not accessible or difficult to copy-paste.
- These materials contain text, images, sound and video, all with or without some levels of tagging, metadata or interactive functionality. Especially in the case of language learning, this functionality can become very complex (Colpaert, 2004). The more ‘enriched’ or interactive these materials, the higher the cost.

On the other hand, language learning content should continuously be updated, adapted, rearranged and rechecked at every change in the learning context. These changes can be due to a curriculum change, a new pedagogical approach such as the 4CD/ID model (Van Merriënboer & Kirschner, 2013) or the Dynamic Systems Approach (Larsen-Freeman, 2009), the integration of a new technology such as tablets or Interactive Whiteboards (Van Laer, Beauchamp & Colpaert, 2012), and to changing learning styles, attitudes and preferences.

Existing learning content is not easy to change. This is mainly due to the fact that most learning content has been created in a dedicated format: it is determined by the medium or the technology of the educational artifact as product. We can distinguish several levels of complexity from low to high:

- Standard multimedia formats such as RTF, DOC, MP3, JPG, and even PDF.
- Enriched standards such as RDF.
- Standards for structuring and transporting data such as XML. While XML is a relatively open and adaptable format, it does not offer enough functionality on the levels of authoring and database queries.
- Dedicated closed database systems with specific structures such as Blackboard exercises.
- Some applications still work with ‘inline data’, meaning that content is provided as simple strings in the source code.

Content gets lost far too quickly due to this inability to adapt, to be reused, exported, transferred. In other words, due to its lack of sustainability.

2. Rationale

In order to remedy this problem, learning content should become more *sustainable*. We define sustainable in this context as the sum of four properties: generic, reusable, interactive and open.

- *Generic*: Content should be authored, structured and accessed independently from any concrete device or medium and should be stored in a separate database. Its structure should not be influenced by any product as possible output.
- *Reusable*: Learning content should be made as transferable or exportable as possible to a wide variety of media, technologies and carriers, such as traditional hard copy textbooks, digital customized printed material on demand, mobile app exercises and materials for Interactive Whiteboard use.
- *Interactive*: Learning content can be ‘flat’ text, audio or video. There are however several possibilities for offering more information (e.g. enriched materials by semantic tagging afford more accurate selection of suitable learning materials) or more functionality (e.g. interactive exercises containing exercise types, answer possibilities, feedback scenarios, error analysis and remediation, reporting and logging).
- *Open*: Learning content should be as accessible, open and authorable as possible for allowing easier co-construction, updating and adaption.

Very little has been published on the topic from a theoretical and pragmatical point of view. Table 1 lists 3 related fields where relevant initiatives have emerged.

Table 1: List of relevant initiatives

Educational modelling	SCORM (Sharable Content Object Reference Model), and later IMS-QTI (Question and Test Interoperability Specification) defined standards for the representation of exercise content and results, supporting the exchange of learning objects between authoring systems, applications and learning environments. Although they can be considered as generic to a certain extent, both standards did not take into account many authoring requirements nor a wide variety of non-electronic output products and services. EML (Educational Modeling Language) was developed by the Open University in the Netherlands; http://celstec.org/content/educational-modelling-language ; EML is no longer under development. EML was taken as a base to develop the IMS Learning Design specification (http://dspace.learningnetworks.org). See page http://edutechwiki.unige.ch/en/Educational_modeling_language. See also Jovanovic et al., 2005; Verbert & Duval, 2008; Manganello et al., 2013.
Content sharing	KlasCement is a Flemish portal for teachers in primary and secondary education. One of its functionalities affords the exchange of learning materials like Hot Potatoes exercises. While KlasCement stimulates exchange and collaboration, it does not support a specific format for doing so. The Flemish PUBELO project (PUBliceren in een Elektronische LeerOmgeving) conducted research on how digital learning content can be made reusable, and on the roles of different actors involved. The project only covered digital content and did not take into account the broader ‘sustainability’ requirements as mentioned above. Knooppunt is a Flemish-Dutch portal for accessing learning content from various publishers with only one registration. The project tackles valuable administrative issues, but does not deal with content structure itself. OER or Open Educational Resources (Colpaert, 2012) is a recent phenomenon

	inspired by Open Source. While OER could reduce teachers' workload and increase learning effect considerably, our current and ongoing research indicates possible technological, psychological and conceptual factors which could prevent this initiative from receiving the impact it deserves. MOOCs or Massive Open Online Courses represent a similar initiative, mainly geared towards opening up existing courses to more participants. The openness of these courses, however, is sometimes questionable. Open Data: this recent approach mainly focuses on making public data available for app developers. Combined with forms of Ambient Intelligence or Augmented Reality, the Open Data phenomenon promising possibilities for education.
Publishing	Most publishing work seems to be done in InDesign, with an eye on HTML5 as strongly emerging standard for digital and more interactive publications. Relevant initiatives are EPUB3 Grid (http://www.idpf.org/epub/30/spec/epub30-overview-20110908.html), and the IWT project Paper to X by VIGC (Vlaams Innovatiecentrum voor Grafische Communicatie).

It should be made very clear that this research tackles a completely different problem than the above mentioned movements and initiatives, but at the same time they are strongly intertwined, complementary and relevant. Content should be stored separately and independently from any electronic learning environment, output product or publishing tool, but at the same time this content structure should take into account these elements. This approach is innovating and novel, yet based on our long track record on the topic.

3. Our track record

3.1. Origins

We have been developing since 1986 a long series of applications and project tools (see detailed CV on www.ua.ac.be/jozef.colpaert > Documents), representing a total of more than 150 man-years in projects for the university, government, institutions, industrial companies, publishers and Europe. Initially the content of these developed programs was stored in a specific dedicated database structure (every application had its own structure). In 1997 we developed a new platform in Windows that focused on two requirements: generate a wide variety of applications with the same source code, and have the content in a separate database. In fact, there were two databases: the first contained all information for the application to run (identity and appearance, menu systems, behavior and interaction) while the second contained the learning content. Both databases were based on different object models, but they were both open, readable and updatable, at least for authorized people. The learning content was stored and structured in a relational Access database. The advantage of this approach was not only the strong integration with Visual Basic programming, but also the fact that authors could easily make their own interfaces as forms, based on queries, and reports. Some authors even succeeded in writing their own error-checking routines in VBA (Visual Basic for Applications).

In the *Eventail/Arcades Interactive Textbooks* project (with Wilfried DECOO), a longer-term project with publisher Van In, we initially converted language textbooks into interactive applications. Gradually, we started first structuring a database of learning content so that CD-ROM and textbook

could be generated at more or less the same time as different output products. We gradually applied this approach in projects where possible: the BIS Online project for the Flemish department of Education, SELOR language tests for the Belgian Civil Service Commission, and a series of European projects. Finally, we ended up working more with the same object models behind the database structure of several different programs. These object models were not technical, but reflected a reasoning for opening, reading, editing and updating a specific database.

3.2. The generic object model

In 2004, we developed an object model that complied with all possible requirements and defended this research as doctoral dissertation (Colpaert, 2004; Colpaert, 2006).

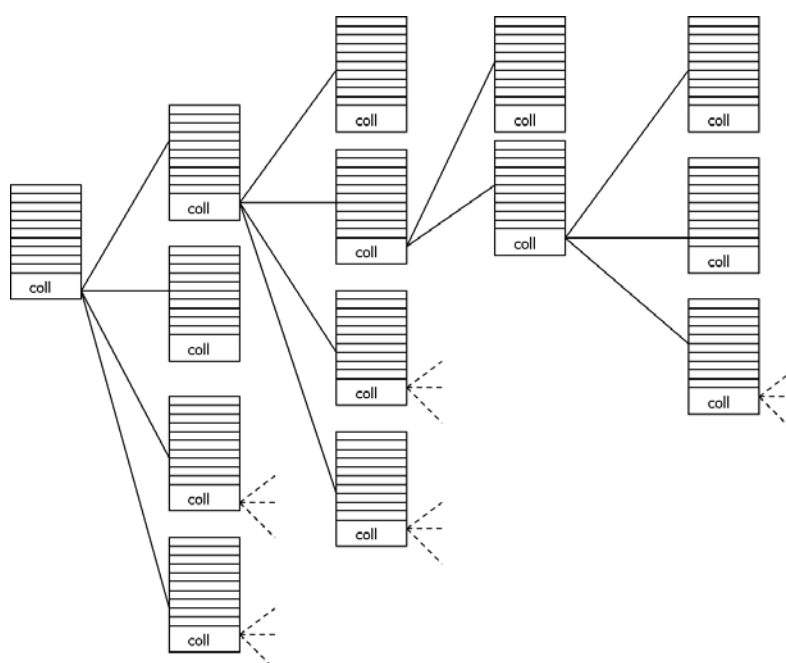


Figure 1. Generic Object Model

The object model (see figure 1) has the following features:

- Simple: the Class of the Generic Object Model only contains a number of properties and methods, the last property being a collection of objects of the same type.
- Scalable complexity: the last Collection property allows the object model to generate itself into a complex tree structure in the case of a long exercise with complex functionality; it can also remain very simple in the case of a plain text.
- The object model pretends to be able to tackle all known exercise, task and interaction types in language learning.
- As a simple object to be called, it represents a powerful communication mechanism between client and server. A complex exercise can be retrieved, loaded, run and returned with only simple lines of code.
- The same object can be used for loading, but also for unloading the exercise, returning information to the server about learner behavior and performance, qualitative data about the learning experience and user feedback.
- The database of learning content can be structured in the same way as this tree-structured software object model. The advantage of this is that a simple recursive routine can then be used to load and unload the object.

Prototyping showed that the proposed object model was compatible with XML (automatic conversion with programming in VB.NET) and also in theory with Learning Object Specifications such as SCORM.

3.3. Subsequent projects

We started focusing more on the database than on software structure. The object model seemed to be applicable to the most diverse topics and functionalities, but the problem appeared to be more psychological and ergonomic.

In a UFOO (University Fund for Innovation of Education) project for our own university, we developed a concrete database structure in Access with interface for developing interactive exercises in Blackboard (Colpaert & Cornillie, 2008). Although this *Entreposage Universel* project succeeded in keeping content outside Blackboard from a technical point of view, it did not deliver the expected results on the users' side. After a thorough analysis of this phenomenon, we found two explanations: a/ the interface was not user-friendly enough and b/ the interface was not generic enough: it was too much influenced by the in-depth structure of Blackboard exercises.

In subsequent projects such as Eurocatering, DISCO and TRUVO, the need for an authoring interface became apparent as content for these projects were being developed in XML/SQL environments. Our research visits worldwide reminded us also that the interface had to be able to generate content to a wide variety of products and services, also non-digital or semi-digital products such as printed materials on demand.

The recently started IWT-project *E-ducate.me* from publisher Diligentia, on the generic structuring of complex learning content for geography, using the same object model for the software. The geography course consists of interactive web materials, but also printed materials.

4. Research description

4.1. The need for a generic interface

While generic structuring (for instance if carried out in a relational database) may appear fairly readable and authorable, surrounding factors in a normal working environment can make things quite complex: large data, many co-authors, complex functionality, author support (queries, forms and reports for content analysis), error checking and prevention (quality control), integration and reuse of existing materials, and generation of content into a wide range of products and services.

In order for the object model to be implemented and to lead to significant and sustainable results, we need to develop an authoring interface, or at least define an ontology. Therefore, two research lines are needed:

- *Technological feasibility*: while the proposed object model has shown to simplify software design to a considerable degree, a number of aspects should be tested thoroughly: testing on specific strategies (graphs, visualization and interaction for mathematics, chemistry and physics), virtual environments, compatibility with SCORM, IMS-QTI and other standards, the possibility to export to DTP programs, security issues, cloud data, OER etc.
- *Psychological aspect*: The generic structuring makes it more complex for authoring at first sight. We need to specify an ontological interface for allowing authors to create content in a supportive, creativity stimulating, functional, reusable and sustainable way.

4.2. Methodology

The adopted methodology is a mixed-method approach consisting of a theoretical, an empirical and an engineering component.

- The theoretical component involves the analysis of multidisciplinary literature and available data in order to yield a complete *status quaestionis* on relevant psychological aspects in authoring (a.o. Cooper, 1997; Colpaert, 2010; Ryan & Deci, 2000; Dörnyei & Ushioda, 2009), and more importantly, to come up with a *conceptual framework* for guiding the next empirical stage.
- The empirical component involves *quantitative* and *qualitative data gathering* about authors and publishers. The quantitative part focuses on how they currently work. The qualitative focuses on how they feel, and what their attitudes and expectations are.
- The engineering part will try to bring together this author's object model, results from the parallel technological research and the results of the previous stages in this project, and generate an ontological specification following a traditional ADDIE design approach.

As a caveat we wish to mention here that it is not our intention to measure any effect of a particular treatment nor to prove that a statement is true or to show correlations. The project does not involve an experimental part as the interface will not yet be used with authors in real-world circumstances. The actual development of the interface and its usage will be the topic of a subsequent project.

5. Expected effects for the educational publishing sector

The above described research should have a positive impact on making development of language learning content an economically more viable and even profitable activity. In this section, we will explain why an authoring interface for sustainable learning content can have a significant effect on education in general, on the production of learning content and on the publishers' business model. Publishing educational content is a vital, but not very viable nor extremely profitable activity in Flanders. Educational publications represent more or less 21 % of all publications in Flanders (for more figures see Delsaerdt 2012).

5.1. Concrete result for educational publishers

An ontological specification should be seen as a detailed description of a system and its behavior. It will consist of a detailed description, visualization, example source code of key routines and possible prototypes of the front end (screen layout and functionality), back end (architecture backbone) and object models.

The granularity (detailedness) of the description should allow any publisher to have the interface developed (and adapted to the local needs and preferences) in a minimum of time and at a very low cost. More importantly, the interface should not be perceived as a technology, but more as a reasoning based on common sense. It should simply feel as the best, safest way of storing their learning content. The concept will be protected under Creative Commons License, and in a subsequent project be distributed as Open Source.

5.2. Expected effect for the sector

Generally speaking, it is difficult to quantify accurately the expected impact of the interface, as all depends on how many publishers will implement it, to which degree and for how many products. Based on the following arguments, as a reasonable **target** to be evaluated, we could state that the development of a unit of learning content would cost 50 % less, and yield 50 % more sales. This is our hypothesis to be validated by comparing expected outcome with realized outcome.

5.2.1. Reducing the production cost

The first expected effect of the interface for sustainable learning content is that it will take less hours for an author to produce a certain amount of language learning content. Thanks to an interface that will offer more possibilities for creating, for querying, reporting, searching, retrieving and

implementing content and to generate the product layer (which makes integrating content in the final product much easier).

5.2.2. Extending the product range

Now almost all learning content is stored in a dedicated structure, determined by a specific medium such as a book or a technology such as an app. Content is mostly linked to one specific output product. The proposed generic model for structuring sustainable learning content allows the output towards a wide variety of products such as ...

Table 2: Product Range

Traditional textbook	The textbook as product remains the same, but content is generated from the generic database into a DTP system for further layout and finishing.
Tailormade textbook	Schools can order tailor-made textbooks, with selected chapters and adapted layout (e.g. with school logo). Printing on Demand (PoD).
Learning content on Demand	Learning Content can be delivered on demand, in any digital format such as PDF, eReader, ...
Learning Content for Specific Purposes	Content can also be delivered in specific formats for specific devices such as Interactive Whiteboards, Presentation Tools (such as Powerpoint or Prezi), Interactive exercises in Blackboard, Hot Potatoes or Moodle.
Open Educational Resources	Publishers can play a new role in contexts where teachers wish to co-construct their own learning/teaching content by providing an interface.
Interactive Applications	Content can be made highly interactive for use in specific applications such as adaptive tests, interactive textbooks, virtual environments and serious games.
Mobile intelligent apps	Limited Interface but huge possibilities in terms of Intelligent Tutoring and Augmented Reality.
Learning Support	Content can be generated into supporting material for e.g. remote human coaching, autonomous learning etc.
Business-to-Business	Learning Content can also be made accessible through an API, allowing other businesses to create Open Data applications.
Researcher data	Finally, publishers can open up their data for commercial and scientific research.

5.2.3. Extending the market

Foreign educational products, especially language methods, are being offered on the market, but they are mostly ill-adapted to the specific requirements of Flemish education. Flemish publishers offer products that are better adapted: written in Dutch, adapted to our pedagogical traditions and educational policy, and generally speaking offering a better didactic quality. This adaptation, however, considerably reduces the market if publishers are unable to quickly adapt their products to another mother tongue or educational setting. And this market is already very small.

The proposed generic interface and database of sustainable content will allow publishers to change the mother tongue or the educational features of a product. Combined with online business technologies, this would open a world market for Flemish publishers.

5.2.4. Sustainability

The separation of content in a generic database, so that this content becomes sustainable on the one hand, and the world market on the other will lead to a new business model.

The new business model entails a new range of products and services, a new production model, new markets, new market pull & market push mechanisms, and above all, a new way of looking at content, learning and technology.

5.2.5. The authoring experience

The targeted interface should lead to more self-efficacy, fun, creativity, motivation and didactic quality for authors. The interface could possibly also be used for teachers who want to involve their students in course creation (such as the author's wiki on <http://ce3.ua.ac.be/wiki>). It could also be used as an interface for creating OERs or with MOOCs.

6. Conclusion

Language learning content can and should be structured and stored in a sustainable way, defined as generic, reusable, authorable (open) and allowing interactivity of scalable complexity. Language learning content can be structured as a collection of a collection of a collection of items, with as many collection levels as deemed necessary: from a plain text (only one level with one collection) to complex task-based scenarios. In this collection structure, it is possible to use the same, simple, object model, consisting of properties and methods which govern appearance and behavior of the object content. The proposed structure appears to work well in client/service environments such as web-based applications or mobile apps, but also for generating a wide range of products and services, also non-digital.

The problem seems to be that it remains quite a challenge to explain to authors how to structure their data in a sustainable way. One of the reasons might be the reason why they should do it, the other being how they would feel empowered and enabled to do so. Psychology will be the crucial element in this respect.

7. References

- Colpaert, J., & Cornillie, F. (2008). Between dedicated and generic requirements: The Entreposage Universel Model as a grammar of interactive content. In A. Gimeno, & R. Seiz Ortiz (Eds.), *Aprendizaje de lenguas asistido por ordenador: herramientas de autor para el desarrollo de cursos a través de la Web* (pp.57-92). Valencia, Spain: Editorial Universidad Politécnica de Valencia.
- Colpaert, J., & Decoo, W. (1999). The role of didactic functions in CALL design. In K.C.Cameron (Ed.), *CALL & the learning community* (pp. 65-74). Exeter, UK: Elm Bank Publications.
- Colpaert, J. (2004). *Design of online interactive language courseware: Conceptualization, specification and prototyping. Research into the impact of linguistic-didactic functionality on software architecture* (Doctoral dissertation). Antwerp, Belgium: University of Antwerp.
- Colpaert, J. (2006). Toward an ontological approach in goal-oriented language courseware design and its implications for technology-independent content structuring. *Computer Assisted Language Learning*, 19, 109-127.
- Colpaert, J. (2012). Open educational resources for language teachers: A goal-oriented approach. In *Proceedings of the joint CMC and teacher education EuroCall SIGS workshop Bologna*, 29-30 March 2012. Retrieved from <http://eurocallsigsbologna.weebly.com/papers.html>

- Colpaert, J.(2010). Elicitation of language learners' personal goals as design concepts. *Innovation in Language Learning and Teaching*, 4, 259-274.
- Cooper, A. (1999). *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. Indianapolis, US: SAMS.
- Delsaerdt, P. (2011). *De winst van de lezer: Inleiding tot het boekenvak in Vlaanderen*. Leuven, Belgium: ACCO.
- Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, language identity and the L2 self*. Bristol, UK: Multilingual Matters.
- Ellis, N.C., & Larsen-Freeman, D. (Eds.). (2009). *Language as a complex adaptive system*. Oxford, UK: Wiley-Blackwell.
- Jovanovic, J., Gasevic, D., Verbert, K., & Duval, E. (2005). Ontology of learning object content structure. In C-K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Supporting learning through intelligent and socially informed technology. Proceedings of the 12th Conference on Artificial Intelligence in Education* (pp. 322-329). Amsterdam, The Netherlands: IOS Press.
- Manganello, F., Falsetti, C., Spalazzi, L., & Leo, T. (2013). PKS: An ontology-based learning construct for lifelong learners. *Educational Technology & Society*, 16, 104–117.
- Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.
- Van Laer, S., Beauchamp, G., & Colpaert, J. (2012). Teacher use of the interactive whiteboards in Flemish secondary education—mapping against a transition framework. *Education and Information Technologies*, 1-15.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning (Second Revised Edition)*. New York, US: Routledge.
- Verbert, K., & Duval, E. (2008). ALOCOM: A generic content model for learning objects. *International Journal for Digital Libraries* 9, 41–63.