# Meaning, Meta Data and E-Learning

By Dave Feasey
Eyepopping Design

March 7, 2002

## Abstract

This white paper analyzes some of the reasons why e-learning has failed, thus far, to deliver on its promise of anytime, anywhere learning. We focus on online self-study courseware as the best example of e-learning content implementation to date and uncover some underlying challenges to adoption of initiatives such as interactivity, a learning- and learner-centered model, learning objects, the emerging e-learning standards and the use of metadata to solve the search and retrieval problem. We recommend a renewed focus on ontologies as machine representations of knowledge domains, and a blended meta data management model as critical parts of an emerging e-learning infrastructure.

## The E-Learning Problem

Online Learning Magazine recently ran a cover story posing the question "Why do learners drop out of online courses? And should it matter if they do?" Completion rates have doubled from 15% to 30%, but this is still not very encouraging. The story suggests a number of reasons for the low figures, including the anectdotal and immediate response that online courses are "long and boring."[1]

### Shovelware

#### Ignoring the Media

The vast majority of online courses are simply repurposed content from other sources- books, classroom course outlines and handouts, CBT- and in most cases the content is not even edited for the online medium by which it's being delivered. E-learning, as it exists today, really is long and boring because it hasn't been designed for its medium. Sometimes the content has been repurposed from classroom handout to CBT to WBT to e-learning without redesign at any of these tranformations.

While e-learning comprises a broad spectrum of approaches, online asynchronous self-study courses make a good example of what's lacking in e-learning today. This type of courseware is the most advanced e-learning content in general.

The e-learning content industry is only a few years old, and the market leaders are those that already had a content presence, namely SmartForce (formerly CBT Systems), Element K (formerly Ziff-Davis Education), NetG (a division of Thomson) and a handful of others. Tool developers have followed content into the market, and to a large extent have reified content's shortcomings into standards models and delivery systems.

A lot of the new content is not new. It has come from previous CBT courses delivered on CD-ROM, courses that were previously derived from book-based curricula. In each iteration the particular strengths and weaknesses of the medium were subordinated to the rush to capture market share in a new category. Those that had existing content were at an advantage if they could get it out of the old form and into the new quickly, that is with minimal editing. The result is essentially classroom books online, sparsely peppered with interactive exercises, multiple choice questions and demonstrations. Jakob Nielsen's usability bible *Designing Web Usability* starts with an expository piece on why people hate to read online, and how much less efficiently they read online, when they do [9]. In fact, web users scan the page for information, rather than reading each sentence of each paragraph. If text is one of the weak points of the online experience, why is it so pervasive?

## The High Cost of Content

Creation of well-designed web courseware is costly for many reasons besides the lost opportunity cost of being late to the party. Instructional Design is itself a highly specialized discipline, and few practitioners are also versed in web design, writing for the web and usability principles, although this is beginning to change. So online courseware has tended to mirror the structure of classroom instruction, that is synchronous / linear, despite the web embodying a diametrically opposed model. The web is asynchronous and excels at networked, nodal, hyperlinked structures.

One of the great promises of online learning has been interactive simulation. The idea of running sales meeting scenarios, playing with virtual constructs and machinery as a learning model is very powerful. Yet the skills necessary for developing such simulations are also highly specialized and expensive. The top authoring tools, Flash and Director attract visual designers most of whom are not trained as programmers, and Java programmers usually are in demand for more mission critical projects than e-learning content development. Very little has been done with DHTML and JavaScript. The call for game-based learning, which can only improve completion rates and retention, runs into even greater upfront cost issues, and so has made little impact on e-learning to date.

Even with a wealth of programmers and design resources, it's very difficult to design simulations to teach soft skills such as sales training, team building or leadership, although it can be done. Roger Schank's *Virtual Learning* suggests hiring professional actors, or better, retired executives to play simulated executive roles, modeled in video clips accessed through complex interfaces [11]. With sales training, for example, the business case for these expensive projects can only be made in companies with huge sales volumes and staffs, where incremental improvements yield sizable returns.

## The Training Model

Leaving aside the expense issue, we see that the design of truly interactive, learner-centric content requires a different Instructional Design attitude, methodology and skill set from that required to develop traditional classroom training. Conceptual material that in the past relied heavily on text can now be communicated in much more direct, visceral ways. One of Roger Schank's great observations (Chris Argyris also makes the point) is that learning occurs through failure [11, 3]. This is sometimes a difficult concept for highly successful, highly educated people who learn by reading. It's not too hard to grasp it conceptually, see that it's correct, but it's difficult to embody and design failure into the work of Instructional Design.

This concept of failure leads to an important conclusion about assessment. Traditionally, we have developed assessment to see how much the student knows at some point in time, either before or after some structured learning event. That is, we've focused on correct answers. The industry is beginning to look at incorrectly answered questions as a way into training content, in other words indicating the material the student still needs to cover. But again, these tests or quizzes are seen as separate from the content, rather than as embedded. Entire course navigation schemes could be built upon frequent periodic assessment. Instead we continue to have static, hierarchical, linear course structures.

We mentioned the networked, nodal, hyperlinked nature of the web medium. Few online courses take full or even partial advantage of this. Hyperlinks perform two functions on a web page. Beside the obvious function of providing a link to another resource, they also stand out from the rest of the text, usually as underlined blue text, providing a scannable digest of the page contents and related topics [9]. Hyperlinks contextualize the page concepts very efficiently when used correctly. The web really is a pull medium, user-centric. Yet taking advantage of this means giving up control of the flow of the presentation. If an Instructional Designer needs to ensure that the student sees certain content in a certain order, she is often in conflict with the underlying model of the web.

## The Learner-centric Ideal

The early e-learning pitch reacted to traditional training by espousing a new, learner-centric model of knowledge acquisition. Marc Rosenberg makes the crucial distinction between e-learning as training (push) and e-learning as information (pull) [10]. This second, learner-centric model implies knowledge management of some sort within an organization (the concept of personal knowledge management is perhaps a bit too new). There are many instances where learners need to be taught or *trained* in something such as sexual harassment policy or OSHA standards. This is the purview of traditional instructor led classroom. On the other hand, many learners need information, for example, on how to merge address data into their Word document, *right now*. They need access any time, anywhere, to this information. Rosenberg identifies this dichotomy as the difference between training and learning [10]. We have been moving to the learner-centric model only slowly, largely due to our concept of learning as something that happens in a structured way or not at all. But there are other reasons as well.

Putting the learner in charge of learning is problematic. First of all, not all learners know how to learn. As Bob Mosher says, we are taught in school how to be learned, rather than how to learn. In his view, part of classroom training involves teaching learners how to learn effectively, since the burden of responsibility is being shifted from instructors to learners [8].

Second, the typically unstructured nature of pull learning is difficult to track. Rosenberg states that content must be delivered to learners in targeted specific ways [10], but this suggests building a system that tracks all user interactions with all forms of learning materials: their books, their software, their phone calls and their conversations around the proverbial water cooler. So far, this personalization of learning resources is non-existent beyond personal portal pages where a student's courses are listed.

Third, we've done a poor job of providing efficient means for learners to locate the needed learning resource. Although algorithms have improved, we still typically talk of "search" engines, rather than "find" engines. The need to locate just the right piece of information in the sea of possible resources is the driving force behind the use of meta data in e-learning. When the computer or other delivery mechanism understands the semantics or meaning of a learner's query and context, then it is conceivable that learning content will be delivered "just in time." Remember too, that in the learning model, resources include internal and external experts, peers, books, help desks and other non-web-based resources. Meta data schemes like Resource Definition Framework (RDF) provide a possible solution by being able to describe non-web resources through unique Uniform

Resource Names (URN). Yet using meta data poses its own conundrums. We'll take up the discussion of meta data later in this paper.

Finally, there are those instances where training is necessary, where competency and completion are required, as in certification programs. Such programs should not be learner-centric.

Actually there is a balance to be struck between training and learning. E-learning, or any other technology, will never fully replace the classroom. Neither will the classroom survive the e-learning challenge unscathed. It will have to change. Our current mental model of learning is heavily biased toward the instructor-centric training model. Up until recently we've used the classroom to serve both training and learning functions. Classroom lectures typically fall into the category of content that can be better and more efficiently presented by electronic means.

### Content is King

The dictum that Content is King is another reason the learner-centric model has not taken off. Content providers have been the winners in the e-learning market so far. As we mentioned, the repurposing of existing, verbose, book-based content onto the web created a vast sea of text content that no one wants to read online. Yet there were significant financial incentives to do this. In fact this sea of content is just what the learning model needs, except that there is typically no "chunking" of the information to allow learners to find the right piece. And content has been webified precisely in order to sell it. If the content is for sale, access to it is restricted at some level and this partially violates Rosenberg's first learner requirement [10]. Vendors package their content and sell to training departments within corporations that make the content available to those that need it. Usually these packages, or packaging mechanisms, are static. The purchaser buys so many seat licenses for IT content, and so many for desktop application content. What happens when learners progress into the IT content? The trade-off between content vendors' needs to profit from their high development costs and the learner's needs to learn are still being worked out.

Content vendors also have a vested interest in channeling learners to their content first rather than to some other resource, even if it might be more effective. Learning Mangement Systems (LMS) need to incorporate better learner control, for example allowing them to tie in their browser's bookmarks, e-mail address books, contact databases, human resource databases, local intranet URLs along with the e-learning content so that a spectrum of resources are presented for a particular query. Below we'll see how machine understanding of semantics and context provides a means for accomplishing these goals.

### The Standards Effort

There are a set of emerging standards in the e-learning world that are designed to realize a vision of e-learning in which there is truly a competitive marketplace for all kinds of content, deliverable as plug-and-play Learning Objects (LO), through any compliant LMS. The concept of the reusable LO comes partly from Object Oriented Programming (OOP) techniques, although LOs are not nearly as useful as programming objects. These LOs are of arbitrary size, but are assumed to be fairly small chunks of content.

Wayne Hodgins, Chair of the IEEE Learning Technlogoy Standards Committee on Learning Object Meta data (IEEE LTSC LOM Chair, for short) states that standardization is a prerequisite for massive growth in an industry. Siting examples such as railroad guages and the VHS/Betamax battle, he shows how these standards preceded hockey-stick type growth curves in these industries [7].

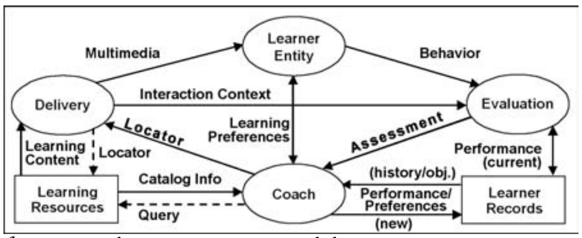

*figure 1. The IEEE LTSA Model*

The standards are designed to encourage the creation of an infrastructure for a vibrant e-learning market. To date the LTSC has largely defined the architecture, meta data fields, Computer Managed Instruction (CMI) methodology, a learner model, and some other pieces to the puzzle, but nothing has been issued as a bonafide standard. Figure 1 shows that this is a balanced approach between a learner entity and a coach. The coach here is assumed to be a machine or system interface, rather than a human being. The coach manages interactions between an LMS, represented on the right side of the diagram, and a Learning Content Management System (LCMS) on the left side. Note that the phrase “learner entity” suggests the learner could be an individual, a group, an organization or even a software agent. So far the coach portion has been built into LMS software as the courseware portal we described earlier. But in the learning model, this coach needs to do much more. It needs to negotiate learning styles and preferences with the learner, and to know about all the learner’s personal knowledge and the organizations knowledge, not just training content that’s been plugged into the LCMS.

## Knowledge Management and E-Learning

The term Knowledge Management (KM) is bandied about in e-learning circles, but as Verna Alee points out, when questioned, the e-learning community doesn’t really know much about KM, Human Capital or Organizational Learning [2]. These are all huge topics, and well outside the scope of this paper, but we need to know a little about KM concepts to make e-learning work [10, 5, 3].

Davenport and Prusak, in their seminal work on KM differentiate between explicit and tacit knowledge. Explicit knowledge is that which takes some physical or electronic form, while tacit knowledge is the internalized knowledge of an individual or group [5]. The expanded diagram of the transfer between these modes (tacit to tacit, tacit to explicit, explicit to tacit, explicit to explicit) suggest a logical progression and a strategy for making knowledge available. The easiest to manage is the explicit to explicit transfer, e.g. from a book to a web page. Tacit to tacit knowledge transfer happens in many ways, but is the most difficult to capture and assess, that is, difficult to make explicit. The learning systems we are talking about deal overtly in explicit knowledge, but need to incorporate the tacit knowledge of the organization’s internal and external expertise [10, 2].

The authors also speak at length about the idea of creating knowledge markets within organizations. They flesh out the model with concepts of exchange and value of knowledge. They also indicate how these markets need to be protected from the same sorts of market pathologies we see in the economy at large: monopolies, supply restrictions, over-regulation and so on [5]. Rosenberg, having written a strategy book, also spends much time on the organizational requirements for effective learning programs [10].

In order to create an effective learning system, we need a way to model our tacit knowledge repositories, the individual and collective expertise of the enterprise. This entails a much larger scope than CBT or online learning has delivered. It needs to encompass tying into ERP and Human Resource systems, for example. While meta data such as RDF can be used to describe or point to individuals, we start to see the breakdown of its use in such systems due to scope, design complexities and maintenance burdens.

## Machine Understanding

The history of machine understanding is its own fascinating story. Artificial Intelligence, Neural Computing, and Natural Language Processing have all aimed to make the Human Computer Interaction (HCI) more humane. All of these disciplines are relevant to building robust learning systems. They are some of the common approaches to modeling and representing tacit or explicit knowledge in an explicit way. Information Retrieval has broad applicability here as well, since we are trying to deliver ever more targeted and specific information by locating it within an ever growing volume of content. One of the simplest and most effective methods of easing this task is the use of meta data.

### Meta data

Meta data is commonly thought of as "data about data" and is commonly used to describe attributes or properties of some piece of information. Frequently, we think of meta data as embedded in the information resource, such as the META tags in HTML documents. But meta data is more useful when it is abstracted out into a separate resource like a library catalog [7]. By describing a resource's properties such as its author, publication date, subject, title and so on, the search engine can reduce the returned set of retrieved items and improve accuracy.

There are two main types or categories of metadata. Resource based meta data describes existing resources by associating property / value pairs to the resource. RDF is such a scheme. Each RDF unit of metadata, a "tuple", defines a resource, a property of that resource, and a value for the property. By using a Uniform Resource Identifier (URI), rather than a URL, RDF gives us the ability to define and describe non-web resources such as people, organizations, locations and even abstract concepts. This is the key descriptive component to Tim Berners-Lee's Semantic Web initiative [4].

Subject based meta data comes at the problem from a different angle. It describes abstract topics, the various relationships between topics, and various explicit resource instances of a topic. XML Topic Maps (XTM) is a scheme to represent such networks of topics and relationships. Ontologies and Entity-Relationship maps can both be described using XTM. Subject based meta data starts to define a way to represent domain knowledge in a more complex and robust way.

The IEEE LOM is a meta data scheme and XML binding used to describe and facilitate location and delivery of Learning Objects (LO). There are over 60 fields or properties described for both LOs and the content components (text, images, audio, etc.) that make them up. The LOM is a resource based meta data scheme.

#### Problems with meta data

There are three main problems with meta data, especially resource based meta data.

First and foremost is the expense of capturing all this data about the already overwhelming body of information available. Not all such meta data can be captured automatically. Thus, early implementations of the LOM have focused on a small, required set of fields, based on the Dublin Core.

The second problem with metadata is that, with its increasing importance, we are likely to experience a meta data explosion mirroring the information explosion it seeks to expiate. Managing the growing body of meta data will soon become a field unto itself. We can only hope such systems are called something other than LMMSs.

Finally, since the meta data schemes we've talked about use URIs to point to resources, and they are external to the resource themselves, anyone can define meta data about anything. Whose meta data do you trust? Some of the issues metadata is supposed to resolve are actually exacerbated by its use.

## Meaning

While resource based meta data adds descriptive properties and values to a resource, there typically is not much meaning or machine understanding implicit in this. Just because I know Shakespeare is the author of Hamlet does not mean I know what an author is, what Shakespeare is or what Hamlet is. In order to start to at least understand the relationships expressed here I need things like Shakespeare is an instance of the author class, an author is someone who writes, Hamlet is an instance of the play class, also a character in the play, the one after whom the play is titled and so on. This larger set of abstract concepts, relationships, contexts and specific instances starts to allow the machine to make more sophisticated judgements, choices and inferrences about knowledge domains. The study of such knowledge domains, at least within the computer field, is called Ontology.

### Ontologies

While a machine-readable representation of a knowledge domain does not necessarily imply understanding, intelligence or consciousness, the effects and results are the same, or close enough. Taking XTM documents as a way of describing an ontology, the machine can merge various Topic Maps to expand its "knowledge." Because ontologies describe relationships and contexts we can request a search engine to return "all plays written by Shakespeare or Marlowe about depressed Danes." And get a list of precisely one play, and a list of resource instances associated with the topic "Hamlet-the-play". The engine knows enough not to look for plays by Phillip Marlowe because that Marlowe is not a playwright. It can also expand the query by searching for additional synonyms of "depressed" in WordNet for example.

While subject and resource based meta data accomplish much the same thing, subject based is much more efficient and robust for some applications within an e-learning context. The use of ontologies can also help ease the resource based meta data problems identified above, by providing a means to automate much of the metadata capture. Since knowledge domains tend to be somewhat static, though complex, investment is better made here since the ontology will be usable across the domain for a significant time period. Describing a resource this way means simply associating it with the correct topic. The list of topics can be greatly narrowed if other meta data about the resource is known to the system.

### Implications for e-learning

The use of ontologies as representations in what is effectively machine understanding creates a more humane, complete and efficient interface between the learner and all the potential resources described in a learning system. It addresses a serious learning content production issue, expands the resource base to include tacit knowledge held by individuals and groups, and may even help reduce overall meta data volume. While ontology design is a specialized skill, and the undertaking is not usually cheap or without its own complexities, cost justifications and business cases need to be made to support ontology use in e-learning systems.

## Conclusion

The next generation of e-learning products will have to do better than the current offerings. Online courseware, though not the only e-learning modality, represents some of what's wrong with current content: it's dull, text-heavy, unengaging, and not designed for the web medium. While there are significant reasons why the first versions of e-learning have been this type of shovelware, and have been based on a training model instead of a learning model, the next stage of e-learning will see diversification of content and models better fitted to particular learning needs and applications, instead of a one-size-fits-all approach. The classroom will also change as a result, as classroom instructors will focus on the strengths of live interaction and move plain old information content out and into e-learning formats.

As we move toward a split model of training and learning for separate needs, many new problems arise including the need to model and track non-web resources and non-computer based interactions, issues of pricing, packaging and business models for content, and teaching learners and instructors how to make the transition to the learner-centric model. We need better search capabilities to streamline location and delivery of diverse sources of learning. As the standards effort comes to fruition, we can expect the content market to be flooded with offerings and the emergence of a new market for meta data management tools.

The current approach to metadata leans heavily on the resource based approach, in which existing resources are defined and "tagged" in various ways for improved retrieval. This approach will need to start incorporating dynamically generated search interfaces that are usable and yield smaller, more targeted search results from an ever expanding pool. Usability will be a key differentiating factor for these search tools.

In addition, a blended meta data approach should emerge, using XML representations of ontologies to ease metadata production bottlenecks and provide even more sophisticated search capabilities. By creating electronic representations of meaning, semantics and knowledge structure, software agents can be employed to streamline the search process further. These knowledge maps are in fact a new type of learning content resource. By focusing on machine representation and understanding as a component of e-learning infrastructure, the e-learning vision of anytime, anywhere learning can come to bear fruit in more productive and innovative learning organizations.

## Bibliography


[1] Alexander, Steve. Do not Pass Go. *Online Learning Magazine* (March 2002), 15-16.

[2] Allee, Verna. ELearning is Not Knowledge Management. *LiNEZINE* (August 2000).

[3] Argyris, Chris, et. al. *On Organizational Learning* (Cambridge, MA: Blackwell Publishers, 1993).

[4] Berners-Lee, Tim. The Semantic Web. *Scientific American* (May 2001).

[5] Davenport, Thomas H. and Laurence Prusak. Working Knowledge (Boston, MA: Harvard Business School Press, 1998).

[6] Guarino, Nicola and Christopher Welty. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM* (February 2002), 45:2, 61-65.

[7] Hodgins, Wayne. The State of Learning Standards. Presentation, OLLO 99, Los Angeles, CA, October, 1999.

[8] Mosher, Bob. How E-Learning Saved the Classroom. Presentation, ASTD E-Learning in 2001, Rochester, NY, February 14, 2002.

[9] Nielsen, Jakob. *Designing Web Usability* (New Riders, 2000).

[10] Rosenberg, Marc. *E-Learning* (McGraw-Hill, 2001).

[11] Schank, Roger. *Virtual Learning* (McGraw-Hill, 1997).