STANDARDIZED FRAMEWORKS FOR DISTRIBUTED LEARNING

Shimon Schocken [1]

The Interdisciplinary Center
P.O.B. 167, Herzliya, Israel 46150
shimon@idc.ac.il

ABSTRACT

Asynchronous Learning Networks (ALN) are often described as "any place, any time" environments for learning [1]. In order for such environments to become feasible and popular on a large scale, they must be based on a public market architecture that is distributed, component-based, non-proprietary, and *standardized*.   This paper describes one such framework, and discusses its possible implementation using a set of standards developed by the *IMS Global Learning Consortium* [2]. In addition, we comment on the central role that XML plays in the implementation of such standards.

KEYWORDS

Asynchronous learning networks, ALN, distance learning, distributed learning, electronic markets for learning, electronic commerce for learning, IMS, online education, universities in the future, virtual education.

I. Background

Since the introduction of personal computers in the early 1980's, instructors and authors worldwide have developed numerous instructional materials in digital form. Most of these materials were designed to support the instruction of specific courses, in specific institutions, for specific audiences. In the absence of a unifying framework, these efforts mushroomed into vast collections of disconnected "knowledge islands," each containing incompatible repositories of lectures, examples, case studies, and problem sets. The introduction of the World Wide Web in the early 1990's has changed the situation, but only on the surface. Scattered subsets of the global inventory of instructional materials are now inter-connected through hypertext links, but the connections are crude and devoid of semantic information.

As a result, discovering and assembling instructional resources that satisfy curricular objectives and methodological requirements is a nearly impossible task. For example, consider an instructor who searches materials to support her next Algebra lecture. Logically, it may be possible to express the instructor's need through a search criterion like (subject= "matrix multiplication" AND context= "economics applications" AND style= "intuitive motivation" AND time<=15 minutes).   Yet presently, there is no Web-based platform from which such a query can be launched and effectively answered. Further, the situation is much worse when it comes to combining materials that come from multiple sources; as Ritter and Suther comment: "it is currently impossible to assemble an Algebra lecture by integrating equation-solving software, graphing skills, and word problems developed by different authors" [3]. The building blocks are out there, in vast numbers and flavors, yet there is no systematic way to assemble them into reusable and inter-operable instructional structures.

To use another example, consider the course that I am presently developing together with Professor Noam Nisan at the Interdisciplinary Center (IDC is a new Israeli academic institution, established in

---

1994). The course, called *Workshop in Digital Systems Construction*, evolves around the construction of a simulated computer, from the ground up. The course is based on a collection of some fifty programming tasks, hundreds of pages of textual explanations and specifications, and various software simulators.[2]   How should we organize these teaching materials on the Web today? The standard solution is to build a series of Web-based lecture modules, whose structure mimics the format in which the course is actually taught in the classroom. This solution will be perfectly appropriate for the students in our course, yet it will completely ignore students in and out of IDC who could benefit from subsets of the course materials, *if they knew how to find them*. For example, students and instructors in *Computer Architecture* courses (in and out of our university) could use the simple *HDL* (*Hardware Definition Language*) simulator that we developed, or perhaps *we* could benefit from such a simulator, if we could *easily* locate it in the digital repositories of similar courses in other universities.

This example illustrates a frustrating phenomenon known to any instructor today. On the one hand, there is an abundance of instructional materials out there. On the other hand, there are no structured ways to discover these materials, inspect their properties, and manage their purchase and usage in an efficient manner. We believe that most of these barriers will disappear with the introduction of *global markets for instructional materials,* as we elaborate in Section 2 of the paper. We describe the goods and services that the markets will offer, and the various market players that will make it work: producers, consumers, and agents. Section 3 explains how this vision can materialize thanks to emerging standards for managing lectures, courses, students, groups, and the related processes of communications and performance assessment.

Throughout the paper I give a series of incomplete and unofficial descriptions of several standards developed by the *IMS Global Learning Consortium*. Readers who wish to obtain the complete standard specifications can downward them from the IMS Web site [2]. Most of these documents, however, are too technical for the general practitioner who simply wants to be informed about the state of the art in the standardization of learning materials. Although I was not one of the authors of the IMS standards, I am using some of their work—in particular the *metadata standard*—to package the instructional materials that I presently develop. With that in mind, one goal of this paper is to give a user-oriented overview of some of the IMS standards, with the hope that other JALN readers will be sufficiently intrigued to experiment with this new and exciting resource.

II. Markets for Instructional Materials

Our notion of a global market for instructional materials is depicted in Figure 1. This section describes the demand, supply, and some of the regulatory forces that will shape the market's behavior.

The goods: At the heart of the market lies a database of autonomous instructional objects, which we call *edupackets*. An edupacket represents an atomic unit of instruction – one that shouldn't be broken further into smaller parts (on pedagogic grounds). The *contents* of the edupacket can be anything: a stack of slides, an annotated picture, a video clip, etc. At the same time, the *package* of the edupacket must be rigidly standardized. In particular, each edupacket is wrapped with a structured set of descriptors that specify its contents using an *agreed-upon metadata protocol*.   Since all the edupackets in the market are characterized by the same metadata, application programs can access and manipulate them via non-proprietary and public software interfaces (like XML). These applications – written by individuals as well as by software vendors -- will facilitate numerous ways to search, edit, and synthesize edupackets, as we describe later in the paper. Importantly, the edupackets database and thus the market itself are assumed to be highly distributed, decentralized, and self-governing. For example, consider a high school teacher who developed an effective way to illustrate

---

[2] A work-in-progress version of the course materials is located in [4].

a certain chemical reaction using a software simulation. If the teacher wants to expose the simulation to the world, she can wrap it up with the agreed-upon metadata and store it in some Web server. The simulation will enter the market automatically, by robot programs that continuously prospect the Web and update the market's inventory (more precisely: inventory of *pointers* to instructional materials located all over the world). Like the Internet at large, the market that we envision will be accessible to all players, and no single player will be able to control it.
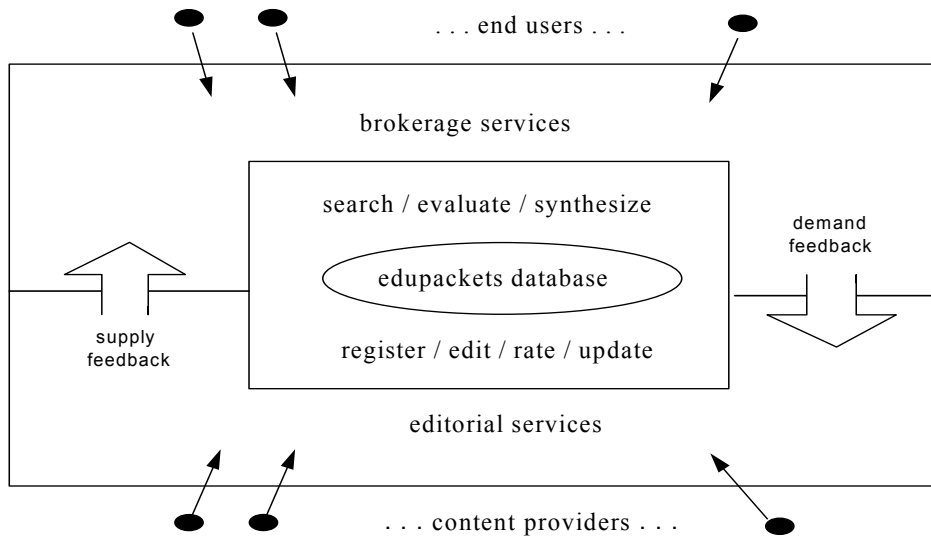

Figure 1: A framework for a global market in instructional materials. The "database" is a logical artifact, representing a highly distributed collection of *edupackets*, scattered across numerous servers worldwide.

Supply side: Principally speaking, any person can contribute edupackets to the market. This includes schoolteachers, textbook writers, corporate trainers, domain experts and retired professionals. The only restriction imposed on content providers pertains to *packaging*: to qualify as an edupacket, the contributed material must be market-compliant, meaning that it must be registered using a data entry procedure that fills in the required standardized metadata descriptors. The content provider has a strong incentive to cooperate with the registration process: the more accurate and complete the edupacket's description, the more likely it will be discovered and used by prospective clients.

The market will most likely be divided into two categories of edupackets: moderated and non-moderated. Moderated edupackets will be edited and priced by publishing agents, and will be offered to the public on a for-profit basis. Non-moderated edupackets will be available as is, on a "use at your risk" basis. This dichotomy is already emerging on the Internet. For example, the "*Educational Objects Economy*" site [5] -- a clearinghouse for educational Java applets -- tags each applet with an optional "*has been reviewed*" label. This feature helps prospective adopters obtain some succinct information about the applet before they bother to test it. In general, the subject of *contents rating* is deliberated extensively in the *World Wide Web Consortium* (*W3C*), and several proposals are emerging. Meanwhile, electronic commerce sites like *Amazon* and *Ebay* employ a variety of proprietary solutions, used to give prospective buyers "market-based" feedback about products and sellers.

Demand Side: The market that we foresee will serve students and instructors from all levels of schooling, as well as corporate trainers and trainees. In fact, any person who has access to the Internet

is a potential market player. For example, people who want to learn more about certain concepts encountered in a Web/TV program should be able to access the market and seek relevant explanations and instructional services about the subject matter. In addition to revealing their instructional needs (either explicitly or implicitly), users will be able to specify time and budget constraints, as well as desired levels of accreditation. This information will be passed to brokering agents that will construct (from the edupackets) alternative instructional services designed to meet the user's needs, budget, and span of attention. For example, consider a user who hears something about neural networks in a TV program. If the user so desires, he will be able to instruct his TV set to arrange for him various instructional treatments about neural networks, ranging from a five-minute abstract description to a one-hour lecture by a domain expert to a full-semester course with or without academic credit.

Agents: In addition to end-suppliers and end-consumers, there will be numerous middlemen and market makers. Most of these players will be automatic agents (computer programs) that prospect the market in an attempt to satisfy their master's needs. For example, a student's agent may continuously search the market for information about a specific subject, taking into consideration the student's scholastic abilities, prerequisite knowledge, and evolving user profile. After finding a set of edupackets from which the student's needs can be satisfied, the student's agent will negotiate usage terms with the edupackets owners (also represented by agents) in order to cut the best deal under the student's budget. The edupackets that will jointly satisfy the student's needs may well come from different sources and authors, creating a need for "integration agents." Needless to say, the design of systems that facilitates on-the-fly integration of instructional materials is a major research challenge from a pedagogical perspective.

III. Architectures and Standards

The global market described in the previous section will remain an obscure idea until a critical mass of millions of suppliers and consumers will enter the fray. These people will come from various quarters such as schools, universities, publishing houses, and software companies. Clearly, all these people will not be able to deal with each other and with the market at large unless they adhere to agreed upon *standards* regarding the storage, management, and dissemination of instructional materials.

In order to address this challenge, several groups of researchers and practitioners in Europe and the USA have began to develop various sets of relevant standards. For example, the following groups have made considerable progress thus far:

- Ariadne: *Alliance of Remote Instructional Authoring and Distribution Networks for Europe* [6]
- IEEE *Learning Technology Standards Committee* [7]
- IMS: *the Instructional Management Systems Project* [2]

Fortunately, unlike the embattled history of standard committees in other areas, there is a great deal of collaboration between IMS, Ariadne, and IEEE. The three groups are sensitive to the social responsibility that they shoulder, and collaborate to promote a worldwide acceptance of the developed standards. The objective of all these efforts is to achieve maximal standardization with minimal specifications. Taken together, these specifications are expected to promote the development of repositories and applications designed to:

- Represent and archive instructional resources in a way that lends itself to world-wide search, packaging, and electronic commerce activities;
- Package "low level" instructional objects into "high level" objects like courses and programs in both academic and corporate settings;
- Support the management of questions, homework and tests;

- Support the monitoring of student activities;
- Support the management of student profiles and the groups to which they belong;

The proposed standards must:

- Be based on publicly available and non-proprietary communications and database protocols;
- Rely, wherever possible, on existing standards in adjacent areas like publishing, electronic commerce, security, and data interchange;
- Provide clean interfaces to institutional enterprise and legacy systems;
- Be widely and internationally accepted by schools, universities, publishers, software companies, and corporate training centers.

The IMS Project: The *Instructional Management Systems* Project [2] was launched in 1997 as an initiative of the non-profit EDUCAUSE organization (a merger of the former EDUCOM and CAUSE associations). In short order, IMS became an international consortium of academic, commercial and government organizations. The group's activities are sponsored by leading computer companies (e.g. *Apple, Cisco, Microsoft, Oracle, PeopleSoft, Sun*), major universities (e.g. *California State U., George Mason U., U. of California, U. of Maryland, U. of Michigan),* and government agencies (e.g. *National Institute of Standards and Technology, US Dept. of Defense, US Dept. of Labor)* [3].

To carry out the standards development work, the IMS has established teams of experts in various areas such as: standard description (metadata) of instructional resources, interfaces with legacy MIS/ERP systems, instructional contents assembly and packaging, questions and tests management, contents management, learner information management, instructional design typology, and accessibility for disabled and disadvantaged learners. The teams proceed in different paces, and their work is reported in three categories of publication: *Supporting documents* contain white and technical papers about the necessary standards. *Scope documents* describe the focus and charter of each standard. *Spec documents* contain standard specifications in various forms of approval (*draft, proposed, approved*). A voting procedure is used to advance standard proposals from one stage to another. We now turn to an informal description of some of these standards and related work.

A. Instructional Resources

The holy grail of a global learning system is a universal mechanism for storing, searching, and synthesizing instructional resources. In order to accomplish this goal, all parties involved, i.e. authors, publishers, resellers, instructors, students, administrators, and software agents, must speak the same language. Unfortunately, the Web's lingua franca– HTML– is a universal language only in terms of contents *rendering*. This is a far cry from the rich semantic structures needed to support a global market for instructional materials.

The lack of semantic transparency on the Web is a well-known problem, and there are several major attempts to "bring order" to Web contents. The most significant development to date is *XML,* or *Extended Mark Up Language* [8]. The upshot of XML is that Web contents should be not only *efficiently rendered* (as is done quite well via HTML), but also *effectively understood* (by humans and machines alike). For example, an XML document containing lecture materials can include not only the materials proper, but also a set of self-describing *attributes* that specify the lecture's author, format, educational objective, etc. Taken together, this attribute set is called the document's *metadata.*

---

3 Only representative members are listed in each category.

XML in a Nutshell: Using XML, an organization (e.g. a company) can design a master metadata template that specifies the minimal metadata that must characterize every document created in the company. This template is called the *document type declaration*, or DTD. Each sub-unit in the company (e.g. the marketing department) can derive from this root DTD a departmental template. As a rule, each template inherits all the attributes of its parent template; at the same time, the template's designer is allowed to *extend* the parent template by adding new attributes to its base attribute set. In execution time, an XML-powered browser (or application) attempts to render (or process) the document using its document template as a road map. The document's template can contain references to other templates, via a variety of techniques called "extensions" and "namespaces" in the XML jargon.

Since XML endows Web contents with *structured* metadata, XML documents can be persisted in corporate databases. The translation of the tree-oriented structure of XML documents to a relational database schema is straightforward. As a result, each document's metadata can be represented by a (possibly complex) tuple of attribute values, along with a link to the document's contents. Using this setting, the organization can archive and manage its Web documents on a database, benefiting from the considerable power of modern database management systems.

XML is rapidly emerging as "the next generation HTML," and sooner or later all organizations will switch to it in order to benefit from the expected avalanche of XML-powered browsers and application programs. From a software engineering standpoint, XML represents a powerful compromise. Its underlying architecture borrows heavily from related ideas in distributed object-oriented design, yet it manages to stir away from the complexities of CORBA and DCOM. Its database persistence model, which is based on periodical reconstruction of new and changed contents, is more robust than the present practice of real-time construction of Web pages from database records. In short, although it is far from being a pure and elegant formalism (and perhaps because of these deficiencies), XML simply *works*.

The IMS Metadata: An IMS group led by Tom Wason developed an elaborate metadata standard, designed for labeling instructional resources of any type, shape and form. The standard is based on a "super template" designed to describe the most general class of instructional resources. The *base metadata* of this super-template consists of 19 core descriptors (e.g. *title, cost, keywords*) and 67 non-core descriptors (e.g. *interactivity level*, *difficulty level*, *intended user's age range*, etc.). The distinction between core and non-core descriptors is pragmatic, and based on surveying prospective searchers and catalogers of instructional materials in the USA, Europe, Asia, and Australia. Quite simply, core descriptors are felt to be more significant (resolution-wise) than non-core descriptors.

The IMS metadata standard is described in four key publications: *Tutorial* [9], *Formal Specification* [10], *XML Binding Specification* [11], and *Best Practices and Implementation Guide* [12]. Table 1 lists the core metadata descriptors, which are discussed in the remainder of this section.

Category	Descriptors
General	Identifier, title, catalog entry, language, description
Life cycle	version, contributor (role, entity, date)
Metametadata	Metadata scheme, language
Technical	format, location
Rights	cost, copyright and other restrictions, description
Relation	kind, resource
Classification	purpose, description, keywords
Educational (non-core)	interactivity type, learning resource type, interactivity level, semantic density, intended end user role, learning

	context, typical age range, difficulty, typical learning time, description

Table 1: The IMS core metadata. The *Educational* category is listed here as a sample of the remaining 67 non-core descriptors. As we explain later in the paper, the *actual* use of IMS metadata requires only a *subset* of these descriptors. Also, when an author labels a particular instructional resource, the values of many descriptors are automatically bound to default values drawn from the organization's profile and from the author's profile. Also, resources that are labeled in batch share many identical descriptor values. In short, prospective adopters of the IMS metadata standard need not be alarmed from the formidable number of IMS descriptors – they are all optional, as we explain later in the paper.

General descriptors: The *identifier* is the resource's GUID (global unique identifier such a URL, URN, or ISBN). *Title, language,* and *description* are self-explanatory. *Catalog entry* is a list of pairs *(x,y)* where *x* is a catalog name (e.g. LCSH = Library of Congress Subject Headings) and *y* is a particular entry in that catalog. This data structure enables parallel indexing in multiple catalogs.

Life cycle descriptors: The *version* descriptor enables the author to register multiple versions of the same resource. This is a useful feature, as instructors and students often need to access old versions of certain resources. The *contribute* family of descriptors identifies the various people associated with the resource's life cycle. These include such roles as *author, publisher,* and *steward*. *Entity* and *date* are, respectively, the person's GUID and the last date in which he or she touched the resource in the respective role capacity.

Metametadata descriptors: It is assumed that in addition to the standard IMS metadata, various organizations will wish to index their resources according to an organization-specific metadata. For example, it is quite likely that a specific university would like to record the *courses* in which the resource comes to play. This *courses* information is relevant to students of that university, yet irrelevant to outside users. With that in mind, the *metametadata* family of descriptors describes the organization-specific metadata (if there is one). *Identifier* and *MetadataScheme* identify the organization-specific metadata scheme, whereas *language* is the technical language in which the metadata scheme is written (most likely, XML).

Technical descriptors: provide basic information about the resource's technical *format* (e.g. HTML, XML, PDF, etc.) and *location* (at present, URL).

Rights descriptors: These answer various questions about the resource's intellectual properties such as: Is it associated with a usage *cost* (which may be zero)? Are there any special copyright restrictions? Other non-core descriptors in that category identify the person(s) with whom these rights can be negotiated?

Classification descriptors: *Purpose* and *description* provide free-text descriptions of the pedagogical objective of the resource. *Keywords* is a list of pairs *(x,y)* where *x* is a taxonomy name and *y* is a keyword identifier within that taxonomy.

Non-core descriptors: The 67 remaining non-core descriptors cover a wide spectrum of instructional properties. Two important usage descriptors are *objectives* and *prerequisites*, each being a list of pairs *(x,y)* where *x* is the name of a domain-specific curricular taxonomy and *y* is the identifier of a specific knowledge unit or skill code within that taxonomy. Using this syntax, *prerequisite* and *objectives* specify, respectively, the ex-ante and ex-post knowledge that the learner possesses before and after successful usage of the resource. The *approach* descriptor specifies the pedagogical methods underlying the resource (*inductive, deductive, procedural, exploratory, emmersive,* etc.). The *User's role* indicates the intended user profile of the resource (*student, instructor, administrator, course*

designer, etc.), and *use level* refers to the user's intended academic level, i.e. school grade, college year, graduate level, etc. *Duration* describes the typical required time span for working with that resource, and *interactivity level* draws its values from the value set {*high, medium, low, none*}.

Discussion: Needless to say, there is a huge gap between the IMS ideal of resource labeling and the prevalent organization of instructional resources, which are by and large unlabeled. Presently, some resources are described by proprietary descriptors (e.g. company-specific labeling of corporate training materials), others by public yet non-IMS metadata schemes, whereas most resources are not labeled at all. How can we start a process of massive labeling of existing instructional resources with IMS metadata? Clearly, requiring contributors to enter 86 descriptor values (for every individual resource!) is not a good way to get started. Labeling with 19 core descriptors "only" is significantly less time consuming, but still too prohibitive.

With that in mind, the IMS compliance team has wisely refrained from specifying which descriptors are mandatory and which are optional. In order for an instructional resource to be "IMS compliant," its metadata must contain *at least one descriptor* from the IMS metadata. Although it is hoped that eventually most resources will be characterized by several IMS descriptors, one descriptor is all that it takes to enter the "IMS fold." For example, consider a hypothetical IMS-compliant search engine, called *IMSFind*, designed to index and locate instructional resources according to the standard IMS descriptors. Suppose that you've labeled an instructional resource that you've authored using a metadata subset containing three descriptors only: *title, description,* and *course-id*. The first two descriptors are part of the IMS base metadata; the third descriptor makes sense only in your own university. This is perfectly acceptable: the *IMSFind* search engine will index your resource according to its *title* and *description* attributes. Users of *IMSFind* will not be able to know the *cost* of your resource (because you haven't specified it in your metadata). Likewise, your non-IMS *course-id* descriptor will be simply ignored by the search engine.

In addition to its liberal conformance policy, the IMS metadata is designed to serve as a "super template" that describes the most general features of *any* instructional resource. Using this base definition as a point of departure, different instructional designers can define different sub-templates of instructional resources, as needed, e.g. *lecture metadata, simulation metadata, case-study metadata*, etc. Any one of these complex types will inherit the base IMS metadata, and will extend it with an additional, type-specific descriptors set. This architecture lends itself nicely to XML: the base metadata can be implemented as the root DTD, sub-templates (specialized resource types) as extensions of the root DTD, and specific resources as individual XML documents ("instances" of the DTD's).

We expect the evolution of new instructional sub-templates to proceed on several dimensions. On the *typological* dimension, generic resource templates will be defined for lectures, simulations, examples, case studies, etc. On the *organizational* dimension, different schools and publishers will create organization-specific metadata, to record organization-specific information. On the *pedagogical* dimension, different authors may wish to characterize their materials according to educational taxonomies of their choice. For example, the educational psychologist David Perkins argues that the design of instructional activities should be guided by four basic questions: (a) What is the *context* (background or purpose) of the subject matter? (b) What is its *structure*? (c) What are some *model cases* of it? and (d) what are the *arguments* that explain and evaluate it? [13]. If a certain organization wishes to characterize instructional resources according to the Perkins typology, it can extend the base metadata with a new attribute called *Perkins type*, and bind it to the values set {*context, structure, example, argument*}.

The key point here is *flexibility:* no one is forced to use a fixed metadata. Every organization is welcome to create its own metadata, provided that (a) the new metadata *extends*, rather than rewrites, the base metadata, (b) the new metadata is implemented in a non-proprietary format like XML, and (c) the organization's metadata template is accessible via a public interface.

Domain taxonomies: The notion of *controlled vocabularies* is crucial for the development of a metadata specification. Wherever possible, descriptors like *subject*, *keywords*, and *prerequisites* should draw their values from widely used taxonomies like NIFL, ERIC, and LCSH. In addition to these general-purpose taxonomies, academic associations and industry task forces are developing domain-specific taxonomies of their own. Using these taxonomies, authors and users will be able to utilize the same terminology when storing and discovering, respectively, instructional resources.

The effort to create domain-specific taxonomies is well underway. For example, the "*ACM/IEEE-CS Joint Computing Curriculum*" [14] divides the Computer Science field into *knowledge units*, each characterized by a symbolic code and a descriptive title. Similarly, many mature professions like medicine, mathematics, and psychology, use agreed-upon knowledge taxonomies. The same phenomenon occurs in the business sector, where different companies, industries and in some cases even countries, have developed internal taxonomies for describing professional skill sets. In a global market for instructional resources, skill codes can be used as a uniform means for describing course pre-requisites knowledge, course objectives, employee and student profiles, etc.

We expect that if an author (or user) will specify a term that does not appear in certain taxonomy, an indexing agent will propose a close match. If no match has been found, the indexing agent will submit a request to extend the taxonomy's vocabulary. This request will be referred to a human editor, assigned to maintain the taxonomy in question. Versions of this "organic growth" of knowledge repositories are already emerging on the Web. For examples, when one fails to locate a desired term in the online "Tech Encyclopedia" site [15], one can request that the term be added to the encyclopedia, and even propose a term definition.

The IMS views the adoption of publicly available domain taxonomies as crucial to the project's success. It is hoped that ultimately, each metadata descriptor will draw its possible values from domain taxonomy or a controlled vocabulary. The *IMS Learning Resource Meta-data Best Practices and Implementation Guide* document [12] lists many recommended taxonomies that can be presently used in the IMS labeling process.

B. From Resources to Courses to Programs

The interaction between users and instructional resources does not occur in a vacuum. Typically, the interaction takes place in the context of a certain course, taken in a certain program, offered by a certain organization. Hence, in order to facilitate their usage, resources must be grouped into high-level instructional units that meet diverse instructional needs. For example, consider an instructional unit on, say, "Mendel's Laws of Genetics." Such a unit can include a variety of instructional resources, as follows:

- Lecture about scientific knowledge in Mendel's time
- Video clip showing how Mendel carried out his experiments
- Multimedia program enabling students to record Mendel's results
- Lecture about the law of large numbers
- Animated simulation of the law of large numbers
- Introductory lecture about genes
- Picture of chromosomes
- Quiz on Mendel's laws
- Quiz on the law of large numbers
- Group exercise in which the students are asked to come up with alternative (non-genetic) explanations to Mendel's experimental results
- Document describing a semester-long project replicating Mendel's experiments
- Spreadsheet for recording the results of the semester-long experiment

- ▪ Test on Mendel's laws

We assume that every one of the above components is an individual resource, associated with its own metadata. We further assume that someone -- an author, an instructor, or a publisher – may want to combine these components into a cohesive package, designed to support a module of instruction about Mendel's Laws.

To address this challenge, the IMS has developed a *contents packaging* standard [16]. A package is a collection of instructional resources that has its own metadata. The package metadata includes various usage-oriented elements such as *title* and *table of contents*, along with a list of all the resources that make up the package, and their *organization*. All this information is represented in the package's metadata, implemented in XML.

We note in passing that the preparation of an instructional package is a problematic issue, and questions abound. How to ensure consistent teaching styles? How to bridge notational differences? How to deal with fractional royalties? And how to price the whole thing? We offer two unsatisfactory answers to these questions. First, it is safe to assume that in most cases, the *core* of the package's components (the lectures, at least) will be created by the same person, who may well be the *packer* herself. Second, the elaborate metadata of the underlying components should help locate and aggregate mutually agreeable resources. In closing, the attractive notion of "let's pack some resources together" obscures some serious problems, yet to be resolved.

C. Questions and Tests

Every instructional process engages the student in many forms of self-study activities: questions, experiments, projects, etc. To minimize clutter, we hereafter refer to all these activities as *questions*. Of course, every practicing instructor has numerous reusable questions in her disposal. First, there is the scattered collection of all the questions that the instructor has prepared in her career. Then there are all the questions prepared by colleagues who teach similar courses, as well as the questions that accompany the course's textbook. Finally, there are question banks that can be either bought or downloaded freely from the Internet.

Yet this wealth of material is largely useless, for three reasons. Since the questions have no metadata, the instructor is unable to locate desired questions according to search criteria. Further, the questions come in an overwhelming hodgepodge of styles and formats, and thus it is difficult to integrate them into a cohesive test or problem set. Finally, most publicly available questions – especially those of the "question bank" stock -- are of a very low quality.

We believe that there is a remedy to most of these problems, in the form of the IMS *Questions and Tests Standard* [17]. The standard is an elaborate specification, designed to structure – among many other question and test authoring and management services -- uniform means for representing, archiving, and combining question and test resources. Following the object-oriented methodology that characterizes the IMS work, each question is viewed as an independent object, wrapped with its own base metadata. The base metadata is viewed as a point of departure, from which many specific question and test templates (sub-types) can be defined.

Presently, the specific question sub-templates include "traditional" assessment tools such as multiple choice, true/false, fill-in-the-blanks, and open-ended questions, as well as "visual" assessment tools such as item-selection, item matching, item-arrangements, etc. Although the present standard captures a wide variety of question types, is also seeks to serve as a basis for stamping out *new* question types, as they become available. Hence, the strategy is to provide a base metadata that can evolve to model the new question types that instructors will want and that technology will permit.

D. People and Groups

In addition to standardized treatment of instructional resources and assessment activities, an integrated system that supports global learning must be based on standard means for representing and managing learner profiles and the groups to which they belong.

People: The need for standardization of learner profiles arises from the observation that learners tend to roam from one learning environment to another. Specifically, people gain knowledge and develop skills in multiple settings including colleges, professional schools, and corporate training programs. Every one of these organizations manages a separate record for the same learner, leading to vast duplication and potential inconsistency. Further, there is no way to "press a button" and obtain a complete record of all the certificates and diplomas that a given student possesses. Also, there is no way to assess more detailed qualifications such as "the student's understanding of algorithms satisfies the formal requirements of the IEEE-97 CS curriculum's AL-15 knowledge unit." Yet in an ALN-oriented world, such information is critical in order to assess the learner's ability to handle edupackets that require pre-requisite knowledge.

In short, what is called for is an elaborate user profile, or *student template,* which can be implemented as an XML data structure. Along with the obvious personal information, the template should include information about acquired skills, degrees, diplomas, and all previous grade transcripts. In short, the student template must carry his or her entire educational and professional history, as well as current professional interests and future plans. To prevent tempering with sensitive information, private fields like *grades* and *diplomas* should be protected and authenticated by the digital signatures of the granting institutions.

To address these needs, Farance and Schoening developed a standard proposal called *PAPI*, for *Public and Private Information*. They describe this effort as follows: "*a key feature of the PAPI specification is the logical division, separate security, and separate administration of several types of learner information: (1) personal information. e.g., name, address, social security number; (2) preference information, e.g., useful and unusable I/O devices, learning styles, physical limitations; (3) performance information, e.g., grades, interim reports, log books; and (4) portfolio information, e.g., accomplishments and works. The PAPI specification may be integrated with other systems, protocols, formats, and technologies.*" [18]

Groups: The communications between instructors and students typically begins at the group level. Contact with a specific student typically occurs when the group-oriented communications fails to satisfy the needs of the individual student. This is particularly true in an ALN-oriented world, where a single course (or instructional resource) may serve hundreds and thousands of learners simultaneously

Indeed, the notion of *groups* is central in any learning process. There are several types of groups, including *explicit groups* (as in all the learners registered to a certain course), *implicit groups* (all the learners that happen to interact with a certain learning resource at a given point of time), and *ad-hoc groups* (all the learners who passed a certain milestone in a certain instructional resource).

With that in mind, future ALN frameworks should be able to continuously create, update, and delete group structures and group membership transactions. In particular, at any given point of time, a vast number of *implicit* and *ad-hoc* groups will be managed as side effects of learner activities. This subject is taken up in the next section, which deals with communications issues.

E. Communications

Education is a communications-rich activity, coming in two main flavors: *synchronous* and *asynchronous*.

Asynchronous communications: A well-constructed instructional resource should be designed in a way that enables an external observer to track the student's performance. One way to implement this

objective is through a set of interactive *checkpoints*, scattered strategically along the resource. For example, consider a 40-hours course, consisting of stacks of lecture slides annotated with synchronized audio presentations. After every group of slides representing a sub-topic, the lecture's author may place a checkpoint that asks the student to report his comprehension of the subject matter, to take a quiz, to ask a clarification question, etc. The lecture's instructor, who may not be the same as the lecture's author, should be able to edit, delete, and add new checkpoints of her own.

Now, we assume that every interaction of student s with checkpoint c is assigned a human monitor t, leading to the basic relationship (s,c,t). Typically, when a group of students $s_1,\ldots,s_n$ takes a certain course, the same teaching assistant t will be assigned to track their performance on all the course's checkpoints $c_1 \ldots c_m$. Yet the basic relationship (s,c,t) is more general, as it enables the course administrator to split the workload between several monitors, assign certain monitors to certain checkpoints and certain student groups, etc. This flexibility will be a fundamental requirement in a virtual course taken by thousands of students around the world.

Before "launching" a certain instructional resource, the course administrator will use a tracking management program to allocate monitors to the various checkpoints associated with that resource. When a certain instructor or teaching assistant t is assigned to monitor a certain checkpoint c, a virtual communication channel (c,t) is established between the two entities. Hence, in most cases, a single person will have to monitor multiple channels. For example, the same teaching assistant may be required to answer questions about this week's lecture, to monitor the student's progress on last week's quiz, and to answer ongoing questions about the course's project. This implies (at least) three independent channels, each connected to an independent *listener* – a computer program designed to collect, process, and present incoming messages to the channel's monitor.

When a student s interacts with checkpoint c, he generates a response $r=f(c,s)$. The result of this interaction (r) is then sent via the channel *(c,t)* to the dedicated listener. More accurately: numerous r values are sent to the same listener, one per each student who passed checkpoint c. The processing of the listener depends on the nature of r. For example, if r is a quiz grade, the listener may be programmed to continuously compute and display the average, variance, and grades distribution of the students who passed this checkpoint, or the names of the students who failed the quiz, or the names of the students who haven't yet taken it, or all of the above. Using this information, the monitor can easily track the student's performance and single out problematic students (or weaknesses in the underlying material), where necessary.

On the other extreme, r may be a free-form question (posed by the student) about a certain aspect of the lecture. Here, the listener can be programmed to (a) parse the student's question, (b) match it on a database of frequently asked questioned (posed previously by the thousands of students who passed *this specific checkpoint* in the past), (c) send a pre-stored answer to the student, (d) encourage him to repeat the question in case the matched answer is not satisfactory, and (e) route the question to a teaching assistant for "human processing." Using this technique, the human monitor will be relieved from the tedium of mundane questions, and will be able to focus on the few questions that merit expert (read: non-FAQ) knowledge.

Synchronous communications: Suppose that a student gets stuck in checkpoint c and needs help. Using a polling program, the student should be able to chat with all the other students in the world who happen to be located in or around checkpoint c in that particular point of time (that is, the students who care to answer).

As opposed to the messages that were described in the previous section, this scenario demands real-time communications. In other words, the sender and the receiver of the message must be connected to the same channel *in the same time*. There exist many other instructional scenarios that demand synchronized communications: multi-student simulations, group interaction, application sharing, and real-time Q&A in online lectures.

The standard: The communications scenarios described above will be based on novel software products, yet to be developed. Examples include channel listening programs, lecture and course management tools, grade reporting systems, help desks, etc. These systems will require various measures of data security, user authentication, and online policing of exam sessions. There is no doubt that all these challenges will be met with effective technology solutions, which are already emerging. Unfortunately, the existing tools are based on proprietary messaging and non-standard interface technologies. As a result, the various tools don't talk to each other, and each ends up addressing a small niche in the student's world.

To cope with this challenge, Ken Schweller of the IMS *Contents Development Team* has proposed a uniform messaging standard, motivated by the following vision: "*a fully implemented Learning Management System will depend heavily on the exchange of data and messages between cooperating applications in a distributed network environment. These exchanges should be secure, durable and often atomic. The applications themselves will vary in their protocols and their abilities to handle synchronous and asynchronous messaging. What is needed is a messaging API that permits these diverse applications to exchange information in a reliable and timely manner without each having to understand all the protocols of the other. A Messaging Service with a well known API makes this possible. The availability of such a service would provide content developers with a reliable messaging infrastructure on which to build their data exchange and communication services, relieving them of having to create proprietary communications protocols.*" [19]

Importantly, the proposed API should be interoperable with existing schemes for distributed object-oriented messaging such as DCOM, CORBA, and RMI. Using this API, software vendors will be able to develop interoperable solutions to support the diverse communications scenarios described in this section.

IV. Conclusion

The paper had two purposes. First, we described a market-based framework for distributed learning in which instructional materials are traded worldwide and combined into flexible course and program structures. Second, we argued that the IMS standards and XML are likely to turn this vision into reality – in our opinion is less than a decade. We believe that standardized markets for distributed learning will change the face of education as we know it, and will enable new models of instruction better and cheaper than those that we presently use. The impact on traditional schools and universities, as well as on third world countries, is likely to be of historic proportions.

V. References

1. **Bourne, J.R., McMaster, E., Rieger, J., & Campbell, J.O.,** "Paradigms for On-Line Learning," JALN Volume 1, Issue 2, August 1997: http://www.aln.org/alnweb/journal/issue2/assee.htm
2. IMS Global Learning Consortium main site: http://www.imsproject.org
3. **Ritter S., and Suther, D.,** "Technical Standards in Education", Educational Object Economy (EOE) Papers Library: http://www.eoe.org/
4. **Nisan N., and Schocken, S.,** "The Digital Core": http://www.idc.ac.il/digitalCore
5. Educational Objects Economy (EOE) Java Applets Library: http://www.eoe.org/FMPro?-db=Categories.fp3&-token=library&-format=/library/JavaApplets.htm&class=Branch&-max=all&-find
6. Ariadne: Alliance of Remote Instructional Authoring and Distribution Networks for Europe Project Site: http://ariadne.unil.ch/
7. IEEE Learning Technology Standards Committee (LTSC)(P1484) Site: http://www.manta.ieee.org/p1484/
8. XML (Extensible Markup Language) W3C Site: http://www.w3.org/XML/
9. Wason, T.D. "Dr. Tom's Meta-Data Guide": http://www.imsproject.org/drtommeta.html

10. IMS Metadata Specification Version 1.1: http://www.imsproject.org/metadata/index.html
11. IMS Learning Resource Meta-data XML Binding Specification:
    http://www.imsproject.org/metadata/mdbind01.html
12. IMS Learning Resource Meta-data Best Practices and Implementation Guide:
    http://www.imsproject.org/metadata/mdbest01.html
13. **Perkins, D.,** "Knowledge as Design", Lawrence Erlbaum Press, 1986.
14. "Computing Curricula 1991", Communications of the ACM, Vol. 34 No. 6, June 1991.
15. Tech Encyclopedia Web site: http://www.techweb.com/encyclopedia/
16. Wason, T.D. "Dr. Tom's Introduction to Content Packaging":
    http://www.imsproject.org/drtomcontentpackaging.html
17. IMS Question & Test Interoperability Information Model Specification Version 1.01:
    http://www.imsproject.org/question/qtinfo101.html
18. **Farance F. and Schoening J.,** "PAPI Specification of Learning Technology: Public and
    Private Information", 1988: http://www.edutool.com/papi/papi-500.html
19. **Schweller, K.,** Supporting Collaboration In an IMS Environment, January, 1999:
    http://ims.collegis.org/collab-
    forum.nsf/ID/25C29C2579E0474E85256713004486B7/$file/CollaborationWP.doc

VI. About the Author

Shimon Schocken is Dean of the School of Computer Science, Interdisciplinary Center (IDC) Herzliya, Israel. Before joining IDC he was associate professor and director of the undergraduate program in information systems at New York University (1986-1996). He holds a Ph.D. in operations and information management from the Wharton School of the University of Pennsylvania. He published in the areas of decision theory, model management, and neural networks, and won the 1987 ICIS Outstanding Dissertation Award. He developed and taught a synchronous distance learning course in database management, and is presently developing an asynchronous Web-based course titled *The Digital Core*.