



IMS Global Learning Consortium

Series on Learning Impact

Access to IMS articles and interviews <http://www.imsglobal.org/articles/>

Taming the Metadata Beast: ILOX

David Massart, European Schoolnet (EUN), Belgium

Elena Shulman, European Schoolnet (EUN), Belgium

Nick Nicholas, Australian National Data Service (ANDS), Australia

Nigel Ward, eResearch Lab, the University of Queensland, Australia

Frédéric Bergeron, TELUQ, Canada

This paper has previously appeared in the November/December 2010 issue of D-Lib Magazine (<http://www.dlib.org>).

Abstract

We propose a framework for organizing multiple metadata specifications in a container that can be handled as a whole. This framework, named Information for Learning Object eXchange (ILOX), is developed as part of the IMS Learning Object Discovery & Exchange (LODE) specification that aims to facilitate the discovery and retrieval of learning objects stored across more than one collection. While thus far ILOX has been demonstrated to resolve a number of challenges specific to the e-learning domain, it is a generic framework that can be profiled to organize metadata about any type of digital content.

Introduction

Learning objects are digital resources used for teaching, learning, or training. Like other types of digital content, metadata (i.e., machine-readable descriptions of learning objects) are used to provide the information necessary to search for learning objects, assess their usefulness, and retrieve them.

There are many ways to look at a learning object. One might be interested in its pedagogical, technical or legal aspects. One might want to know how it is used in practice, how its users perceive it or how accessible it is to people with special needs. All these aspects are important and have to be taken into account in order to efficiently find, retrieve and reuse a learning object. A variety of metadata specifications exist that capture different aspects of learning objects. Some general purpose specifications such as IEEE Learning Object Metadata [1] and Dublin Core education [2] capture the main aspects of learning resources, whereas specialized metadata schemes permit one to produce detailed descriptions of a particular aspect of a learning object (e.g., IMS Accessibility For All [3] and Contextual Attention Metadata [4], which enables descriptions of the accessibility of a resource and its actual usage¹, respectively).

Collecting and organizing all the information available about a learning object is difficult. One can either keep the various metadata elements in separate documents that reference each other or create ad hoc metadata profiles that combine relevant pieces from various specifications. None of these solutions is entirely satisfying. In practice, references between metadata records prove difficult to maintain and process while ad hoc profiles are usually defined within limited communities outside of which they are not interoperable. Moreover, these patchworks are generally difficult to create, understand, and maintain.

The IMS Learning Object Discovery & Exchange (LODE) [5] specification aims to facilitate the discovery and retrieval of learning objects stored across more than one collection. It can be seen as a glue specification that profiles existing general-purpose specifications in order to take into account requirements specific to the educational domain, rather than creating new specifications. Among other things, it proposes a framework, named Information for Learning Object eXchange (ILOX), for organizing multiple metadata specifications in a container that can be handled as a whole.

Note that, although initially addressing problems faced by the e-learning community, ILOX is a generic framework that can be profiled to organize metadata about any type of digital content and is not limited at all to metadata about learning objects.

This paper is an introduction to ILOX. Section 2 describes how the conceptual model of ILOX combines the Functional Requirements for Bibliographic Records (FRBR) data model [6] with a powerful abstraction mechanism named materialization [7] to organize metadata records. Section 3 presents the application profile of ILOX used by the Learning Resource Exchange, a service that allows European teachers to get access to digital educational content from many different countries and providers [10].

¹Alternative initiatives to provide a description framework for usage data include CEN/ISSS work on social data [12] and NSDL work on paradata.

Modeling Learning Objects with FRBR

In information modeling², materialization [7] is used to represent the relationship between a class of categories (e.g., learning objects) and a class of more concrete items (e.g., learning object copies). Materialization is important in formulating metadata for learning objects, because it captures commonalities between descriptions of objects at different levels of generality: metadata attributes may apply at a more abstract level, to a larger number of instances, or at a more concrete level, to a smaller number of instances.

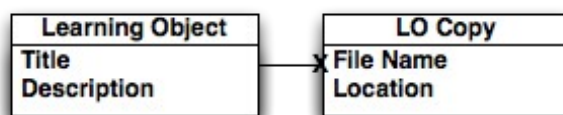


Figure 1 – learning objects have multiple copies.

The class diagram of Figure 1 presents an example of materialization. It relates a more abstract class: Learning Object to a more concrete one: LO Copy. Class Learning Object represents the information about learning objects (e.g., their titles, their descriptions) whereas class LO Copy represents the information about concrete copies of these learning objects (e.g., the file names and path of these LO copies). Materialization is noted as a straight line with a * at its more concrete class end.

The attributes of abstract classes are propagated to the classes materializing them. So if a learning object has the title "Hamlet", then all LO Copies materializing it also have the title "Hamlet". This allows us to express metadata economically: we need only to define title once for a learning object, rather than repeating it for each LO Copy.

Materialization provides a powerful mechanism to structure metadata descriptions. In the bibliographical domain, the Functional Requirements for Bibliographic Records or FRBR [6] can be modeled as a materialization hierarchy that is useful for distinguishing between aspects of learning objects relevant to different contexts. The FRBR concepts of "work", "expression", "manifestation", and "item" as they relate to learning objects are illustrated in Figure 2.

² "Information modeling is concerned with the construction of computer-based symbol structures which capture the meaning of information and organize it in ways that make it understandable and useful to people." [9]

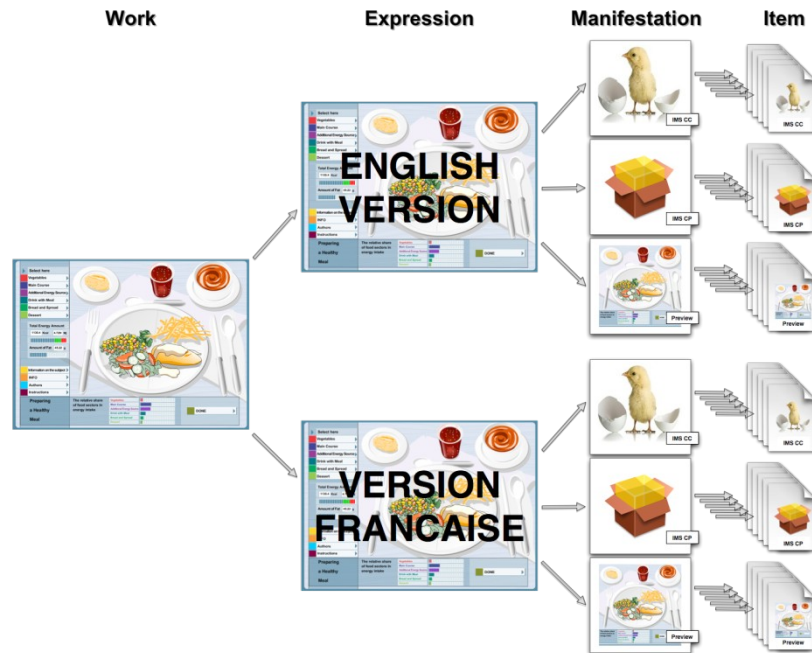


Figure 2 - Example of different FRBR expressions, manifestations, and items of a learning object work.

A FRBR "work" is a distinct (intellectual or artistic) creation, such as the learning object about nutrition shown on the example of Figure 2. Different versions of this learning object can exist: for example, an English and a French version. These versions are different FRBR "expressions" of the work. Each version of the learning object can take different forms. For example, the English version of the learning object about nutrition can be available as a preview, an IMS Content Package and an IMS Common Cartridge. Each of these different embodiments of an expression of a work is referred to as a FRBR "manifestation". Finally, copies of the IMS Common Cartridge of the English version of the learning object may exist in a number of locations. Each of these copies is a FRBR "item".

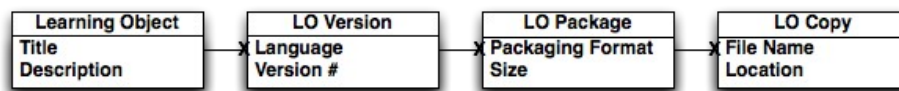


Figure 3 – Using materialization to model the relationship between the different FRBR representations of learning objects.

As depicted in the class diagram of Figure 3, materialization can be used to model the relationships between different FRBR aspects of learning objects. Class LO Copy models the item aspect of learning objects. Class LO Copy is the materialization of class LO Package, which models their manifestation aspects. In turn, class LO Package is the materialization of class LO Version, which models the expression aspects of learning objects. Finally, Class LO Version is the materialization of class Learning Object, which models the work aspect of learning objects. To save space, each of

these four classes in the example is shown with only two attributes typical of the aspect that it models.

The ILOX data model structures a learning object description as a materialization hierarchy such as the one presented in Figure 3. The FRBR materialization levels are used as follows:

- **Work** is an abstract view of a learning object that captures the commonalities between all the possible variations of this learning object such as, for example, the pedagogical content that is common across all the variations of the learning object.
- **Expressions** are used to capture information specific to the different versions, drafts, translations, and localizations of learning objects, such as language.
- **Manifestations** are used to capture information specific to the way a given expression of a learning object is encoded and presented, such as file formats.
- **Items** are used to capture information specific to the concrete copies of learning objects, such as the URI where they can be accessed.

Some types of information will typically be specific to one FRBR materialization level. For instance, the language of an object is typically characteristic of an Expression: an object may be translated into a different language without becoming a different Work, but different Manifestations of the same Expression are all expected to be in the same language. However, other types of information may appear at multiple materialization levels: access rights may apply to all copies of a Work, or may be specific to a particular Manifestation (e.g., a preview of the learning object vs. the runtime object). The same information at a lower materialization level overrides information appearing at a higher level. (So access rights can be set for the Work as a whole, but access rights for a specific Manifestation can be treated as an exception.)

An ILOX instance can be rooted at any level of the hierarchy depending on how abstract or concrete one needs to be. Handling learning object descriptions at the:

- **Work** level permits one entry per learning object with no immediate distinction between learning object versions;
- **Expression** level permits one entry per learning object version with no immediate distinction between the different formats of a given learning object version, and without having to decide which Work different Expressions belong to;
- **Manifestation** level permits one entry per learning object format with no immediate distinction between the different copies of a learning object, and without having to decide which Work or Expression the Manifestations belong to;
- **Item** level permits one entry per learning object copy, without having to decide which Work, Expression or Manifestation the Items belong to.

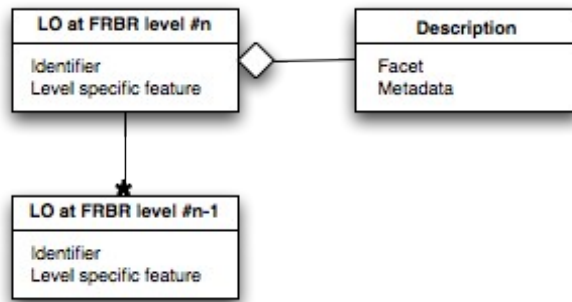


Figure 4 – Pattern for describing the different FRBR aspects of a learning object.

At each level of the hierarchy, a common pattern is used to model the corresponding FRBR aspect of the learning object. This pattern is shown on the class diagram of Figure 4 where a given FRBR level (modeled by class "LO at FRBR level #n" is described by:

- Optional identifiers (modeled by attribute "Identifier"),
- Descriptions consisting of level-specific metadata (modeled by class "Description"),
- Additional level-specific information (modeled by attribute "Level specific features"), and
- Information about the immediate lower FRBR level (modeled by class "LO at FRBR level #n-1").

"Descriptions" are used to describe each FRBR level of a learning object with level-specific metadata. They consist of two components:

- **facet** indicates what "facet" of the given FRBR aspect of the learning object in question is described, and
- **metadata** contains a metadata description of the given FRBR aspect of the learning object in question.

Each level can have multiple metadata descriptions, each with its own facet to differentiate between them. ILOX does not define a controlled vocabulary for facets. Instead, application profiles of LODE are expected to select controlled vocabularies for the facet elements. These vocabularies generally differ from one application profile to another and from one FRBR level to another.

The Learning Resource Exchange and its Metadata Application Profile

The European based Learning Resource Exchange (LRE - <http://lre.eun.org/>) federates metadata from a variety of learning object repositories and provides a service allowing teachers to access the learning objects from various access points. Potentially, any application that utilizes learning objects can connect to it. European Ministries of Education make learning objects accessible for their own teachers via national portals. When learning objects have the potential to 'travel well' for use in contexts beyond their national origin, content providers describe them with metadata using the LRE Metadata Application Profile [10] and expose this metadata so that it can be easily accessed by the LRE. In turn, the LRE compiles the collected metadata to produce a digital catalog of learning resources that can be consulted by teachers using the LRE or their own national portals.

In the LRE, obtaining a learning object is a three-step process:

1. The first step involves discovering and evaluating metadata in order to select a learning object that meets a user's need.
2. The second step is negotiating access to the selected learning object. This step can require authentication, authorization, and encryption schemes depending on the learning object level of protection [11]. For learning objects that are freely available at the specified location, the negotiation step is perfunctory.
3. The third step is retrieving the selected learning object at the location obtained during the second step.

Controlled vocabularies describing the pedagogical qualities of the learning objects such as learning resource type, subject, typical age range and learning contexts (among others), translated into 24 languages, are integrated in the LRE Metadata Application Profile (LRE MAP). The vocabularies are managed and made accessible using a browsable interface and for machine-to-machine processing in the Vocabulary Bank for Education.

Federating sets of metadata coming from various origins, with content provided by ministries of education (MoE), commercial and non-profit content providers, and cultural heritage organizations poses a number of challenges. One of the more pressing needs for the LRE in federating metadata is to overcome the limitations of a reliance on a single metadata specification such as IEEE LOM without undermining interoperability and backward compatibility as needs and requirements continue to evolve. Furthermore, a rapid rise in the production and dissemination of complex learning objects (in multiple languages, in multiple formats, in multiple locations, tailored for particular populations and dedicated platforms) necessitates a more precise way to indicate which aspect of the object is being described in a single metadata record. Finally, the generation of metadata about learning objects is no longer within the strict purview of the objects' creators and trained indexers. The ascent of social networking cultures has created opportunities and expectations that users and networked communities of practice will generate and trust social metadata to guide their choices about services and products; including learning objects. Such user-generated comments, bookmarks and other types of evaluations are producing valuable streams of information for building recommendation systems, structuring search result rankings and feedback

channels for content creators. All participants in the LRE federation (i.e., users, content providers, and portal managers) benefit if such social data can be captured, aggregated and transported in a single metadata container with all relevant available information about a learning object for use in multiple contexts. Current metadata specifications used in the e-learning domain such as Dublin Core and IEEE LOM do not allow for the capture, aggregation and dissemination of social metadata without undermining interoperability. The following scenario describing the types of learning object metadata managed by the LRE vividly demonstrates the challenges of metadata management in the e-learning domain:

A metadata record for a learning object, “Resistance in a Wire” is cataloged by the LRE from a harvest using OAI-PMH protocol. The creators of this learning object licensed it under a Creative Commons license allowing for its reuse and sharing with proper attribution. The object, a simulation allowing users to manipulate a wire’s resistivity along with activities and lesson plans is available in three languages, English, French and Spanish. The learning object’s versions are available in several formats. While the English version can be rendered in a web browser it also comes packaged as an IMS Common Cartridge version 1.1, which has a more restrictive license than the web-based format. The French and Spanish versions are available only as IMS Common Cartridges and access to them resides behind a login wall. The IMS Cartridge format of the English version has been rated and bookmarked by several hundred teachers who have used it. Download statistics have been collected when the object packaged as an IMS Common Cartridge has been downloaded in English, French and Spanish. Knowing the language version of an object may not be enough in the European context with a variety of educational systems and curriculums. The same learning object in French is also tailored for the French educational system. There is also a Swiss system version. Finally, the English version that can be rendered in a web browser also allows for setting to make the object available for visually impaired learners.

The challenge is to describe all this information in one metadata record and provide users with an ability to discover the version and format of this learning object that meets their needs and to evaluate the object’s suitability with the help of recommendation systems.

IMS LOD Information for Learning Object eXchange specification (ILOX) in combination with the IEEE LOM Metadata standard (LOM) [1] has been selected as the basis for the Learning Resource Exchange Metadata Application Profile v4.5 [10] because it can address the evolving requirements of a learning object repository federation by providing for interoperability of metadata, the ability to identify what is being described by metadata and the use of multiple metadata specifications in one metadata record.

As illustrated on Figure 5, the main commonalities shared by all subsequent levels of this learning object will be described with an IEEE LOM metadata instance attached at the ‘main’ facet of the root level (in the LRE MAP, the Work level is the preferred root level). The LOM will include general elements such as title, description and keywords as well as pedagogically relevant elements such as learning resource type, age range for typical users, intended educational context, etc.

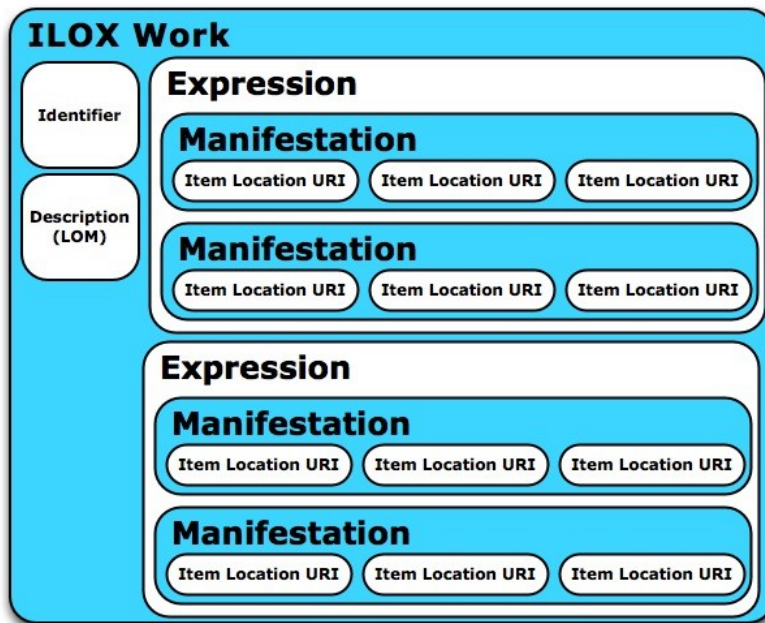


Figure 5 – Describing the “main” commonalities shared by all FRBR levels using IEEE LOM at the ILOX Work level.

A license/rights facet is available for use at this (and every) level. We can use the license/rights facet to attach metadata stipulating the Creative Commons license terms. More restrictive license terms for versions and manifestations at the lower level of the ILOX will have their own license/rights metadata attached and the license stipulations of the more abstract level will be superseded by the license information at the more concrete levels. For the Spanish and French versions available as IMS Common Cartridge with a restrictive license, we would attach the rights information at the Manifestation level.

Furthermore, because the copies of the IMS Common Cartridge format of the French and Spanish versions are behind a login wall, a transaction facet will be attached at the Item level to indicate the steps necessary for negotiating access to the retrievable copy of the object. (To meet these requirements, the LRE has developed an Access Control Metadata Schema that must be attached at the “transaction” extension point [11].)

The ILOX Expression Dimension type provides solutions to make explicit the relationship between the versions, their formats and ultimately the location where these items are available for retrieval. Using the Expression Dimension Type “language” we can indicate that the object is available in French, English and Spanish. We can also use the “coverage” Dimension Type to indicate when the object’s version has been tailored to meet the needs of a particular region, which differentiates it from versions that are in the same language but that are intended for another educational system. For each Expression of the learning object a Manifestation is mandatory. The LRE uses controlled vocabularies for Manifestation names such as “thumbnail”, “experience” (a web page), “preview” as well as “package in” for objects that are packaged in some form. In this scenario the Manifestation name is “package in” (an IMS Common Cartridge v1.1) and “experience” for objects that play directly in the browser (in this case only for the English version of the learning object).

There are two ways to express information about the accessibility features of the learning object the first way is to indicate accessibility as a versioning of the object. Another way is to attach metadata describing the accessibility features of the object at the ILOX Expression level. In the case of the learning object described above, by using the facet mechanism at the Expression level we can indicate that the English version of the object also offers features tailored for visually impaired learners by attaching metadata at the “accessibility” facet describing the options available. This illustrates how ILOX allows for taking advantage of standard schema such as “IMS Access for All” [3] for addressing specific requirements.

The LRE Metadata Application Profile also provides for the use of a ‘reputation’ facet at any level to capture any type of user generated assessments of a learning object (ratings, annotations, bookmarks) that can aid in the object’s retrieval and rankings, work with recommendation systems and/or support social navigation tools. Ratings and comments made about the English version available in the web browser can be attached as metadata at any level and provided to the national portals where the ratings can be used to sort results for most popular learning objects or to let teachers browse objects that have been rated or vetted by fellow teachers. A schema to support these requirements is in development by an expert team of the CEN Workshop on Learning Technologies [12].

Using the paradata facet we can capture and aggregate information produced by recording meaningful actions and processes users initiate to locate and access the learning object in this scenario (e.g., web server logs). Such data includes number of visits, number of downloads, etc. This data is initially collected at the item level and then can be aggregated at different upper ILOX levels using the paradata facet. Such aggregation is intended to track the number of times different formats of different versions of an object were accessed, starting from the number of downloads of individual copies at the Item level. Aggregations will be available at each level for exchange and sharing. For example, the number of downloads of an object can be collected for each item and then aggregated by format (Manifestation), i.e. 10,000 downloads for all Common Cartridge formats and 15,000 times played in a web browser and then aggregated at the Expression level to track version preferences. This information can be offered to the content provider as feedback to understand user preferences in different national contexts and with different formats.

Thus, using ILOX in combination with LOM and other metadata specifications makes it possible to organize all these specifications in one metadata container. Using level specific attributes at the ILOX Expression and Manifestation levels makes it possible to provide information on the ways versions differ from one another and then provides information allowing for the efficient retrieval of those versions in all their available formats and locations. The facet mechanisms of the ILOX allow for social and meaningful actions’ data to follow a learning object through its life cycle. When different versions or formats have special features or specific rights’ stipulations, these can also be effectively expressed all in one metadata container.

Conclusion

This paper presented the IMS LODE Information for Learning Object eXchange (ILOX) data model, a framework for organizing, in a semantically meaningful way, multiple metadata specifications in one container. ILOX is a generic framework that can potentially be used to organize metadata about any kind of resources.

This framework facilitates the collection and handling of the diverse information necessary to efficiently retrieve learning objects. It allows for the processing of all the metadata about a resource as an entirety and for integrating in one container all of the appropriate specifications.

This framework is being developed as part of the Learning Object Discovery & Exchange (LODE) specification [5] of the IMS Global Learning Consortium [13] with the support of the ASPECT project [14] that used ILOX as a basis for producing a new version of the LRE Metadata Application Profile. The latter makes it possible to easily manage the discovery and exchange of learning resources in multiple formats and versions.

It is important to note that because ILOX is a framework for organizing metadata rather than a new metadata specification, it was possible to completely automate the generation of ILOXes from existing metadata records thus easing the adoption of the new LRE application profile by the LRE content providers.

Taking advantage of the experience gained with the LRE metadata application profile, the IMS LODE group is now working on an IMS profile of ILOX for learning objects. We expect this profile to be ready by February 2011.

Acknowledgments

The work presented in this paper is partially supported by the European Community eContent*plus* programme - project ASPECT: Adopting Standards and Specifications for Educational Content (Grant agreement number ECP-2007-EDU-417008). The authors are solely responsible for the content of this paper. It does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of information contained therein.

1 References

- [1] IEEE Standards Department. IEEE 1484.12.1-2002, Learning Object Metadata Standard. July 2002.
- [2] Dublin Core Education. <http://dublincore.org/groups/education/>.
- [3] IMS Access For All v2.0 Final Specification. <http://www.imsglobal.org/accessibility/>.
- [4] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking actual usage: the attention metadata approach. *Journal of Educational Technology & Society*, 10(3):106-121, 2007.
- [5] D. Massart, N. Nicholas, and N. Ward. IMS GLC Learning Object Discovery and Exchange Base Document v1.0. Base document, IMS Global Learning Consortium, March 2010. <http://www.imsglobal.org/lode/>.
- [6] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report, volume 19 of UBCIM publications, new series. K.G. Saur, München, 1998.

[7] A. Pirotte, E. Zimanyi, D. Massart, and T. Yakusheva. Materialization: a powerful and ubiquitous abstraction pattern. In J. Bocca, M. Jarke, and C. Zaniolo, editors, Proc. of the 20th Int. Conf. on Very Large Data Bases, VLDB'94, pages 630-641, Santiago, Chile, 1994. Morgan Kaufmann.

[8] D. Massart, "Towards a pan-European learning resource exchange infrastructure". In Y. Feldman, D. Kraft, and T. Kuflik, editors, Proceedings of the 7th conference on Next Generation Information Technologies and Systems (NGITS'2009), LNCS 5831, pages 121-132, Haifa, Israel, June 2009, Springer.

[9] J. Mylopoulos. Information modeling in the time of the revolution. Information Systems, 23(3-4):127-155, 1998.

[10] D. Massart, E. Shulman, and F. Van Assche. Learning Resource Exchange Metadata Application Profile version 4.5. European Schoolnet, 2010. <http://lre.eun.org/node/6>.

[11] J.N. Colin, T.D. Le, and D. Massart. A federated authorization service for bridging learning object distribution models. In M. Spaniol, Q. Li, R. Klamma, and R.W.H. Lau, editors, Advances in Web-Based Learning – ICWL 2009, LNCS 5686, pages 116-125, Aachen, Germany, August 2009. Springer.

[12] CEN WS-LT Social Data. <https://sites.google.com/site/censocialdata/home>.

[13] IMS Global Learning Consortium. <http://www.imsglobal.org/>.

[14] Adopting Standards and Specifications for Educational Content (ASPECT). <http://aspect-project.org/>.

About IMS GLC

From Innovation to Impact

The mission of the IMS Global Learning Consortium is to advance technology that can affordably scale and improve educational participation and attainment. To ensure that the “Learning Impact” of technology-enabled innovation is achieved around the world, IMS's influential community of educational institutions, suppliers, and government organizations develops open interoperability standards, supports adoption with technical services, and encourages adoption through programs that highlight effective practices. For more information visit www.imsglobal.org or contact info@imsglobal.org

Citation

To reference this article please cite:

Massart, D., Shulman E., Nicholas, N., Ward, N., Bergeron, F. (2010). Taming the Metadata Beast: ILOX. IMS Global Learning Consortium Series on Learning Impact. November 2010 from <http://www.imsglobal.org/articles/nov2010ilox.cfm>

Copyright © 2010 by Massart, D., Shulman E., Nicholas, N., Ward, N., Bergeron, F.

All Rights Reserved.

The IMS Logo is a registered trademark of IMS GLC.

tel: +1 407.362.7783

fax: +1 407.333.1365

<http://www.imsglobal.org/articles/>