

# Creating a unified digital content supply network for global education: the Global Grid for Learning story so far

Theodore Lynn<sup>i</sup>, Samuel DiGangi<sup>ii</sup>, and David Hanney<sup>iii</sup>

<sup>i</sup> The Learning, Innovation and Knowledge Research Centre, Dublin City University

<sup>ii</sup> The Applied Learning Technologies Institute, Arizona State University

<sup>iii</sup> Global Grid for Learning, Cambridge University Press

**Abstract.** Throughout the world significant investments have and are being made in the design and development of digital learning content repositories and brokerage systems to facilitate the exchange of metadata and associated content to K12 educational software systems. While attempts have been made to include commercial content providers, most of these initiatives are organized regionally and financed by government, sub-government or donor agency funding and focus on the sharing of open education resources. The Global Grid for Learning is an initiative funded by Cambridge University Press, a not-for-profit organization, with a vision to create a unified digital content supply network for global education connecting users and software systems to both commercial and open education resources. This paper provides an overview of the initiative and discusses some of the content and metadata challenges encountered to date and how this project has sought to address them. The paper also outlines current efforts made towards systems interoperability.

**Keywords:** digital content, brokerage systems, global education, metadata, interoperability

## 1 Introduction

Cambridge University Press [1] was founded by a royal charter granted to the University of Cambridge by King Henry VIII in 1534. It is the oldest printer and publisher in the world, having been operating continuously since 1584, and is one of the largest academic publishers globally. Cambridge University Press is a not-for-profit organization governed by the Press Syndicate, which consists of eighteen senior members of the University of Cambridge who oversee the Press's business. The purpose of Cambridge University Press is to further the University's objective of advancing learning, knowledge and research.

In line with the Press's mission and its role as a global publisher, Cambridge University Press participates in numerous research, publishing and philanthropic

projects worldwide.<sup>1</sup> Many of these projects have involved the delivery of content through digital learning repositories in an effort to provide end users with universal, affordable and equitable access to digital learning resources however few had global scope, self-sustainable funding models, the required investment in infrastructure and support resources to satisfy modern service level agreements or sought meaningful engagement with the commercial publishing industry. While Cambridge University Press, like many publishers, believe in a digital future, the financial and other costs to participate in that future are considered by some to be speculative and in some instances prohibitive, particularly for local and smaller niche publishers.

Following further consultation with education stakeholders and architecture design work with Atomic Assets Limited and Microsoft, the Global Grid for Learning initiative was launched by Cambridge University Press in March 2007 [2]. The vision for Global Grid for Learning is create the digital content supply network for education worldwide and to provide universal, affordable and equitable access to digital content, whether commercial or community developed. The goal is to connect a billion digital resources to education worldwide by 2018.

It comprises two services, one for commercial content providers which is managed on an operational basis by Cambridge University Press, and one for community content providers which is managed by a number of academic partners, namely the Learning Innovation and Knowledge Research Centre (LINK) at Dublin City University [3], the Applied Learning Technologies Institute (alt<sup>^</sup>I) at Arizona State University [4], and the Centre for Applied Research in Educational Technologies (CARET) at the University of Cambridge [5]. Both services leverage the same underlying platform to create an ecosystem of consumer-providers of commercial and community content and to:

- enable software manufacturers, independent software vendors, portals and ultimately end users to seamlessly and dynamically discover and access up to date learning assets and structured learning objects hosted or offered by one or more content providers;
- enable content providers to offer subscription services or free access to their learning assets or structured learning objects to software manufacturers, independent software vendors, portals and ultimately end users.

The commercial service (GGFL.COM) was launched in November 2007 in the USA and Canada and will be launched in the United Kingdom, Ireland and the Arab World in October 2008. The community service (GGFL.ORG) will be launched in September 2008 [6].

---

<sup>1</sup> For example, Cambridge University Press is a participant on the eContentPlus projects - MELT and ASPECT.

## 2 The Global Grid for Learning Model

Global Grid for Learning is an educational broker as defined by Simon, Oberhumer and Kristofl:

“Educational brokers are network-based information systems, in the sense of educational mediators, which integrated learning objects from dispersed sources in order to make them available to a wider audience. Educational brokers can be perceived as enhanced digital libraries, which include computer-mediated communications, community process support and/or marketplace functionality. Educational brokers focus on different kinds of educational artefacts ranging from small learning assets (e.g. a picture of an elephant), to full-fledged online-courses (e.g. A course about the wild life in Africa). In order to integrate and distribute these educational artefacts, educational brokers provide interfaces to learning management systems, local repositories of educational material, course catalogue management systems, assessment tools, etc.” [7]

Content Providers register their services and content to a central directory operated by Global Grid for Learning. Commercial content providers enter into standard license agreement which allows distribution of their content within an aggregated library or as a discrete collection. Content Providers may participate in GGFL.COM and GGFL.ORG. For example, Cambridge University Press is making content available in GGFL.COM but is also making a subset of their content available in GGFL.ORG under a Creative Commons license. In this way, Cambridge University Press hopes to bridge the gap between the commercial and open content provider communities.

Global Grid for Learning negotiates distribution agreements with software manufacturers, independent software vendors, portals and resellers (Market Intermediaries) and provides support for integration of the Global Grid for Learning Web Service with their software systems. Market Intermediaries may be commercial or not-for-profit organizations including inter-institutional consortiums, government or sub-government agencies and projects. Once integrated, the end users of the Market Intermediaries can seamlessly search, discover and select content from Global Grid for Learning suitable for their end user requirements through their existing software interfaces, typically virtual learning environments, content authoring software or internet browsers. Global Grid for Learning is typically not the seller of record and therefore end users typically contract and accept end user license agreements with the Market Intermediary. Except for the GGFL.ORG portal due to launch in September 2008, the end user is typically fully or semi-anonymous to Global Grid for Learning i.e. Global Grid for Learning at best can only identify the institution to which the end user is attached. This was an important consideration to address fears of competition, channel conflict and the use of user information for direct marketing activities.

The Market Intermediary may be an open or community content initiative and wishes to access open or community content through Global Grid for Learning. In this instance, Global Grid for Learning attempts as much as possible to encourage two-way sharing so that content availability is maximized. All market intermediaries, whether commercial or not-for-profit, must abide by the Global Grid for Learning End User License Agreement or the license requirements of the individual content provider.

For commercial content in an aggregated library, license fees are negotiated with each Market Intermediary by Global Grid for Learning on behalf of all Content Providers. Content Providers can choose not to participate in a given distribution agreement on a case by case basis or provide standing instructions. License fees are distributed to Content Providers on a periodical basis based on relative usage in the relevant period. This usage-based compensation is critical as it incentivizes Content Providers to contribute more content, higher quality content, more relevant content and better metadata. The current method of calculation usage is based on a data transferred. While providing a simple solution, this method has been deemed inadequate as it may inadvertently give a greater weighting to higher resolution images, badly designed content or “bloatware”, and legacy files with less effective compression technology. It has highlighted the need for a content equivalency model which can be used to price or weight learning assets and structured learning objects relative to each other.

Commercial content providers set their own license fees where content is offered as part of a discrete collection. In this instance, Global Grid for Learning may negotiate an agreement with a Market Intermediary on behalf of the Content Provider or may be directed to distribute the content to a Market Intermediary to meet an a priori agreement between the Content Provider and the Market Intermediary. A small standard percentage of net aggregate fees are paid to Global Grid for Learning for service and infrastructure costs; no upfront fees are charged or paid to content providers. No fees are charged or paid for community content distribution.

Global Grid for Learning’s business models are flexible and focused on how Content Providers shape their business models and adjusts its services accordingly such as supporting distributing open or free content, pay-per-download or limitless downloads by annual or monthly subscriptions. As part of the service, all service infrastructure is paid for and managed by Cambridge University Press, who maintain a 24/7/365 technical support team through a best practice service level agreement.

Cambridge University Press uniquely combines a long-life not-for-profit organization with a global commercial presence. This provides the initiative credibility with both commercial Content Providers and the wider education community. This credibility is perceived as essential from a sustainability and service level perspective.

### 3 Content and Metadata

By June 2008, over 1.25m resources were available to education users through GGFL.COM with a further 1.25m resources under license. Over 40 commercial publishers and rights management organizations are currently participating in Global Grid for Learning, representing several thousand sources worldwide. These participants range in size and scope and include some of the leading names in international publishing such as Encyclopaedia Britannica, Reuters, Corbis, Bridgeman Art Library, Cambridge University Press, Cambridge-Hitachi, Scran and others. Content Providers participate on a non-exclusive basis and on the same general terms. Typically, smaller commercial content providers prohibit distribution of content in home territories. Participation is open to all content providers, regardless of size.

Content is required to be appropriate for education use. This has been widely interpreted and Global Grid for Learning has had to introduce guidelines as well as automatic filtering against specific watch-list. This has been a particular issue with large image libraries and is likely to become an increasing issue as the service is introduced to culturally-sensitive regions such as those in the Arab world. For example, the Saudi Arabian Communication and Internet Technology Commission will block any sites that contain content that is considered related to pornography, drugs, alcohol, gambling, terrorism and bomb-making or contains slanders or abuse directed towards the Islamic religion or Saudi laws and regulations [8].

Resources include learning assets and structured learning objects [9]; the former comprising predominantly image, audio, video and text files, and the latter comprising IMS Content Packages. To maintain neutrality within the system and between Content Providers, GGFL.COM does not provide a system for user-generated ratings or evaluation, although usage statistics will be provided. Similarly, to cater for country context, pedagogic practice and indeed language, Global Grid for Learning does not mandate contextual metadata rather it leaves this to the Content Provider and the end user, where possible. Choice is deferred to user control; the end user, whether educator or learner, decides whether a particular piece of content is appropriate and relevant to their needs, not Global Grid for Learning.

To remain neutral and to accommodate as much content and as many devices as possible, Global Grid for Learning does not enforce a particular file or packaging format. All media formats are supported however advanced functionality including automated metadata ingestion, thumbnailing and transcoding are available for a limited subset of file types per asset class, as appropriate. This subset includes:

- Images - PNM, PNG, JPEG, BMP, TIFF, EPS
- Documents - PDF
- Videos - MPEG, AVI, SWF
- Audio - FLAC, MP3, WAV
- Learning Objects - IMS Content package

Content is typically supplied in the highest quality format. CYMK and RGB color profiles are supported.

Content metadata definition is a fundamental part of the architecture as searching content by metadata is a significant functionality of Global Grid for Learning. Unlike content packaging and open education initiatives, the publishing industry has not settled down to a particular standard defining common attributes of learning assets and learning objects. Metadata readiness typically falls in to a number of scenarios:

- (i) Public sector funded content which meets IEEE LOM e.g. content developed under the UK National Learning Network programme
- (ii) Content providers with digital asset management systems and proprietary or segment-specific metadata schema e.g. IMARS, IPTC, ONIX, IIM
- (iii) Content providers who have little or no metadata at the disaggregated level

To participate in the Global Grid for Learning, content providers must provide a minimum set of mandatory and recommended metadata elements per resource similar to those proposed by Simon and Oberhuemer (2006) [10]. The mandatory elements include unique identifier, file name, file type, title, and keywords. Optional elements include subjects, education levels, curriculum data, description, extended description, rights statement, and DPI. Curriculum data may vary in granularity. As such, Global Grid for Learning caters for the following base constructs:

- 1. Level
  - a. Course
    - i. Subject
      - 1. Topic
        - a. Sub-topic

Details on the content providers are collected during the completion of legal agreements. Global Grid for Learning is not reinventing the wheel but merely expressing metadata requirements in language that content providers are more familiar with. Once data is provided it is expressed in more commonly used technical language and is aligned with a metadata schema largely based on IEEE LOM.

Few commercial Content Providers can or will provide access to their systems. This has caused some unexpected issues. For example, the logistics of transferring high quality images or video in bulk from one system to another is significant and in some instances had not been attempted by the Content Providers with external parties. This can result in significant workload and system load for all parties involved particularly where one is dealing with hundreds of thousands of files and several terabytes of data. In addition, files may contain metadata unknown or unused by the content provider e.g. GPS data. Typically, once the files and metadata are received they are ingested in to the Global Grid for Learning system as collections on a content provider and subject basis within Global Grid for Learning's system.

Similar to the process outlined by Prause et al (2007) in relation to the MACE initiative [11], metadata is enriched in three ways:

- (a) Automatic enrichment: where available, existing metadata is extracted and in some cases enriched without human interaction. Where metadata exists in a file, Global Grid for Learning automatically extracts metadata depending on file-type e.g. keywords, title, catalog, catalog entry, description, IMS organization (for content package viewing), pixel dimensions and colour space. In addition, non-file-type specific data is generated on ingest e.g. file byte length. Visualizers have been implemented for various file-types to generate thumbnails and preview images. An adaptor object converts any visualizer into a metadata extractor that extracts pixel dimensions. This, for example, is how the pixel size of a video is identified. Specific custom metadata extraction code for IMS Content Packages and JPEG images has been developed. All of this is handled by a configurable MediaTypeStore where each file type is configured thereby and facilitating the extension of metadata extraction capabilities on an ongoing basis.
- (b) Semi-automatic enrichment: Once ingested, the Global Grid for Learning content team supervise the enrichment of metadata using a combination of semi-automated rule-based tools. The enrichment typically relates to adding and cross-walking metadata relating to national curricula and education standards, digital rights metadata and in the case of commercial content, licensing and pricing data may be added. For example, following user feedback, a series of rules have been generated to facilitate the discovery of assets that have a high match probability with specific educational standards and calendar dates (for pushing to end users on specific dates). Content specialists review these results and meta-tag them to appropriate standards, rejecting inappropriate suggestions.
- (c) Manual enrichment: some resources are not easily tagged and require additional input from educators with specific domain knowledge.

In the case of only one content provider to date, metadata information is periodically harvested from the content provider's base repository using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In this instance, once content is selected by the end user, it is downloaded from the Content Provider's system through Global Grid for Learning. This has caused unexpected problems in commercial implementations as occasionally there are differences in the metadata stored by Global Grid for Learning and the Content Provider, which only becomes apparent at the point of download by an end user e.g. asset file size where asset file size is a determinant of price. Furthermore, the likelihood of failed content retrieval is increased where dependencies on third party service levels are introduced. While this may not be a major issue for open content initiatives, it impacts directly on commercial service level agreements and in some instances, may introduce financial penalties.

As part of the MELT and ASPECT projects, Global Grid for Learning will support Simple Query Interface (SQI). Initially, this will be used to share community content on GGFL.ORG on the European Schoolnet Learning Resource Exchange and possibly other similar initiatives such as GLOBE. Surprisingly, a number of open content initiatives, while enthusiastic about receiving content from GGFL.ORG do not wish to share content available through their initiative in bulk or through a live interface with GGFL.ORG, even though it may be available under Creative Commons.

Publishers of educational content are required and in some cases, mandated, to correlate content to specific education standards by Departments of Education to qualify for funding (e.g. US No Child Left Behind Act or UK eLearning Credits programme) or to meet discovery and usability expectations. This can be a costly exercise:

- In some markets, there are different national, regional and local standards i.e. in the US, there are over 75,000 federal, state and district standards which a publisher might have to meet
- Different terminology exists in different markets causing semantic interoperability issues e.g. “Key Stage 2” in the UK, “manipulative” in Canada etc
- Different languages may be used in different markets e.g. British vs American English, Arabic etc
- Where content is disaggregated, each discrete object must be correlated i.e. instead of a CD-ROM of 100,000 images having one correlation, 100,000 may be required
- Each asset may be correlated to multiple standards in one curriculum i.e. an image of global warming could be correlated to a science standard, a geography standard and a civic studies standard
- Standards are often not available as machine readable datasets and are subject to change regularly e.g. in Ireland only the primary curriculum is available in machine-readable form
- Content is replaced and/or upgraded regularly requiring new correlations

Existing solutions are either very US-focused or entail a high element of manual tagging and therefore are extremely costly. This is perceived as a significant barrier for smaller Content Providers who wish to sell their content in international markets, and particularly in the US.



To help address this issue, Global Crosswalk, a joint project between LINK, altVI and the Global Grid for Learning was established. The project explores how curriculum and education standards between different regional and national jurisdictions impact on the discovery and usage of digital content in the schools sector worldwide. The project is developing tools to produce high quality alignments between metadata schema and specifically from (a) industry or segment specific metadata schemas to educational metadata schemas and vice-versa and (b) from one curricula or set of educational standards to another and vice-versa. For example, a content provider completes a boilerplate profile. During ingestion, metadata is extracted and stored to a master metadata schema. Fields in this master metadata schema contain the content provider's boilerplate data and have been aligned with other schema including Dublin Core, IEEE LOM and ONIX for Books. The end user can access different views (e.g. Dublin Core, IEEE LOM etc), complete missing fields using drop-down boxes. The initial work concerns the design and development of a US vocabulary against which all major US education standards will be crosswalked. All correlations will be capable of delivery as data streams and reports easy-to-use formats such as XML, Word and Excel. Current evaluations of the project are resulting in significant cost savings resulting from reductions in discrete multi-State correlations.

#### **4 Systems Interoperability**

Systems interoperability is a critical success factor for Global Grid for Learning. Similar to Facebook with Thrift, Global Grid for Learning has designed an API that allows the technical team to define data types and service interfaces in a simple definition file. Taking that file as input, the compiler generates code to be used to easily build RPC clients and servers that communicate seamlessly across C#, Java or PHP. The API provides download and metadata discovery services and is optimized to allow many operations in a single round trip thereby decreasing the effect of network latency.

The API has been implemented in SMART Learning Marketplace [12], a Global Grid for Learning implementation with SMART Technologies currently live in North America. SMART Learning Marketplace is also available within the SMART Notebook software, a leading interactive whiteboard and lesson creation tool. In addition, standard web parts exist for SharePoint 2003 and MOSS 2007 to support virtual learning environments using Microsoft SharePoint. Global Grid for Learning is currently working with Arizona State University and Dublin City University on integration paths for Sakai and Moodle.

#### **5 Conclusion**

Global Grid for Learning unifies and integrates appropriate technologies necessary to provide Content Providers, whether commercial or community-based, and Market

Intermediaries in the education sector with a system, repository, and enabling workflow process for distributing learning assets and structured learning objects. Global Grid for Learning enriches metadata to enhance the likelihood of discovery and the relevancy to the end user. The initiative promotes interoperability, open standards and collaboration within and between the commercial publishing and open content communities. To this end Global Grid for Learning supports the exchange of learning assets and structured learning objects through other, possibly competing, initiatives.

## References

1. Cambridge University Press, <http://www.cambridge.org>
2. Global Grid for Learning, <http://www.globalgridforlearning.com>
3. DCU Learning, Innovation and Knowledge Research Centre, <http://www.dcu.ie/dcubs/link/>
4. ASU Applied Learning Technology Institute, <http://alti.asu.edu>
5. University of Cambridge Centre for Applied Research in Educational Technologies, <http://www.caret.cam.ac.uk>
6. GGFL.ORG, <http://www.globalgridforlearning.org>
7. Simon, B., Oberhuemer, P. and Kristofl, R.: The Everlasting Dawn of Educational Brokers – A Search for Key Design Principles. Paper presented at the First International Workshop on Learning Object Discovery and Exchange, 18 September 2007.
8. Communications and Information Technology Commission: Content Filtering in Saudi Arabia, [http://www.internet.gov.sa/learn-the-web/guides/content-filtering-in-saudi-arabia/view?set\\_language=en](http://www.internet.gov.sa/learn-the-web/guides/content-filtering-in-saudi-arabia/view?set_language=en)
9. McCormick, R.: Keeping the Pedagogy out of Learning Objects. Paper presented to the Symposium Designing Virtual Learning Material, EARLI 10th Biennial Conference, Improving Learning, Fostering the Will to Learn, 26-30, September 2003.
10. Simon, B. and Oberhuemer, P.: Austrian Learning Network for Higher Education – Metadata and Interface Specification for Interoperable Educational Systems. [http://www.fnm-austria.at/strategiecontent/Dateiablage/download/Metadadata\\_Model\\_2006-06-26\\_Version\\_1.1.pdf?file%5fid=23005](http://www.fnm-austria.at/strategiecontent/Dateiablage/download/Metadadata_Model_2006-06-26_Version_1.1.pdf?file%5fid=23005), FNMA, 2006 as cited in Simon, B., Oberhuemer, P. and Kristofl, R.: The Everlasting Dawn of Educational Brokers – A Search for Key Design Principles. Paper presented at the First International Workshop on Learning Object Discovery and Exchange, 18 September 2007.
11. Prause, C.R., Ternier, S., de Jong, T., Apelt, S., Scholten, M., Wolpers, M., Eisenhauer, M., Vandeputte, B., Specht, M. and Duval, E.: Unifying Learning Object Repositories in MACE. Paper presented at the First International Workshop on Learning Object Discovery and Exchange, 18 September 2007.
12. Smart Learning Marketplace, <http://learningmarketplace.smarttech.com>.