Learning Resource Metadata Initiative: using schema.org to describe open educational resources.

Phil Barker, LMRI project, Cetis & School of Mathematical and Computer Sciences, Heriot-Watt University.
phil.barker@hw.ac.uk
Lorna M Campbell, LMRI project, Cetis, Institute for Educational Cybernetics, University of Bolton.
lorna.m.campbell@icloud.com

Abstract
This paper discusses the Learning Resource Metadata Initiative (LRMI), an international project that aims to facilitate the discovery of educational resources through the use of embedded metadata that can be used by search engines (e.g. Google, Yahoo, Bing, Yandex) to refine the search services they offer. LRMI has extended the schema.org metadata vocabulary with terms that are specifically relevant to aiding the discovery of learning resources. In order to understand LRMI metadata it is necessary to first introduce schema.org metadata. LRMI was later funded to assist a number of services providing open educational resources in modifying their display interfaces so that they included relevant embedded metadata. In addition to synthesizing the lessons learned from these implementations, the current phase of the LRMI project is engaging with potential users, including search providers and learning resource developers, with the aim of extending the uptake and implementation of the specification.

Keywords
Metadata, resource discovery, learning resources, LRMI, schema.org

Introduction
Since June 2011 the Learning Resource Metadata Initiative project (LRMI) has been working to make it easier for teachers and learners to find educational materials through major search engines and specialized resource discovery services. The approach has been to extend the schema.org ontology so that educationally significant characteristics and relationships can be expressed. In this paper we will introduce the technology approach taken by LRMI, describe some example implementations and demonstrate how the discoverability of open learning resources may be enhanced. In our presentation we shall outline the background to LRMI, including its aims, who is involved in achieving them, and technical aspects of the specification. We shall present the latest developments in LRMI metadata implementation, drawing on lessons learnt through analysis of its use by publishers of open educational resources and the learning registry, and will report on any developments regarding its adoption by major search engines.

In order to understand the approach LRMI has adopted, it is first necessary to understand schema.org, so we start with an introduction to schema.org before focusing on the specific contribution of LRMI to open educational resource discovery.

Schema.org
Schema.org was launched in June 2011; it is a joint initiative undertaken by Google, Yahoo, Microsoft Bing, Yandex and W3C (schema.org, n.d.). The aim is to improve the display of search results, making it easier for people to find the right web pages. Content publishers insert machine readable information into the HTML of web pages which helps search engines

understand the significance of the text on those pages. For example: without such markup a web page might contain the following HTML.

```
<h1>Learning Resource Metadata Initiative: using schema.org to
describe open educational resources</h1>
<p>Phil Barker, Cetis, School of Mathematical and Computer
Sciences, Heriot-Watt University</p>
<p>Lorna M Campbell, Cetis, Institute for Educational Cybernetics,
University of Bolton</p>
```

A search engine might be able to guess that the content in the `<h1>` tags could be used as the title of the resource (though often even this will not be obvious), but it may struggle to understand the nature of the resource being described (i.e. that it is a conference paper), the identity of the authors and which text represents the authors' names and which relates to their affiliations.

Schema.org has two components; an agreed ontology, i.e. a hierarchy of resource types and a vocabulary for naming the characteristics of resources, their relationships with each other, and constraints on how to describe these characteristics and relationships; and the expression of this information in machine readable formats such as microdata and RDFa Lite.

For the example above, the resource as a whole is a Scholarly Article, the title is the name of the Scholarly Article, the authors are People, who are in turn affiliated to certain Organizations. Expressed as microdata embedded in the HTML of the resource, the snippet above would become:

```
<div itemscope itemtype="http://schema.org/ScholarlyArticle">
<h1 itemprop="name">Learning Resource Metadata Initiative: using
schema.org to describe open educational resources</h1>
<p itemprop="author" itemscope itemtype="http://schema.org/Person">
   <span itemprop="name">Phil Barker</span>,
   <span itemprop="affiliation">Cetis, School of Mathematical and
Computer Sciences, Heriot-Watt University</span></p>
<p itemprop="author" itemscope itemtype="http://schema.org/Person">
   <span itemprop="name">Lorna M Campbell</span>,
   <span itemprop="affiliation">Cetis, Institute for Educational
Cybernetics, University of Bolton</span></p>
</div>
```

The whole of the HTML describing the paper is wrapped in a div element that identifies it as a single item of type `ScholarlyArticle`. The h1 element gets an attribute to specify that it contains the `name` of the item (i.e. the title of the scholarly article), which is a simple text value. The `p` elements get attributes to specify that they each contain a `description` of distinct items that are each a `Person` who is an `author` of the scholarly article; some of this description is marked with a `span` element to show that it is the `name` of the `Person`, some is marked to show that it represents an `affiliation` of the `Person`.
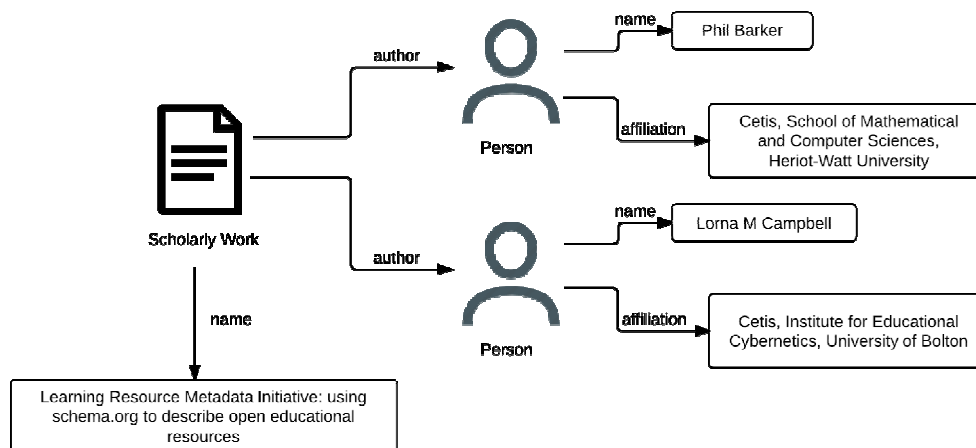
Figure 1, showing some of the relationships and properties expressed in a description of a scholarly work.

A search engine can parse this data to extract the graph of information shown in figure 1. More sophisticated markup is possible, for example providing unique identifiers for the all the items, and marking up the authors' affiliations as well-described Organizations rather than just providing a text string. Even more significant are the other article properties that can be added, e.g. description, subject, publication date, publisher, copyright holder, cost, etc. This information, if extracted by a search engine, can be used to enhance the search interface in order to facilitate resource discovery. An example of such an enhanced search interface for shoes is shown to the right. For scholarly works one could imagine a search interface that allowed you to filter by a specific author or subject, or to specify that the resource must be free.

LRMI: Learning Resource Metadata in schema.org

Many of the important characteristics of a learning resource can be described using the properties schema.org uses to describe generic resources. For example the author, publisher, ISBN, publication date, availability, cost etc. of a textbook can be described in the same way as any other type of book. However there is some information relating to the educational usefulness of a resource that is not covered in the original

Figure 2: Results from a Google shopping search showing examples of how further refinements can be made on properties such as those which may be marked up using schema.org.

schema.org ontology. As a result, the Learning Resource Metadata Initiative was established to identify additions to the schema.org ontology that would make the discovery of learning resources easier (LRMI, n.d.).

LRMI is led by Creative Commons and the Association of Educational Publishers (and so covers both open education and commercial publishers of learning resources) with funding from the Bill and Melinda Gates Foundation and the William and Flora Hewlett Foundation. It was established in June 2011 and is now in its third phase of funding.

As a result the phase 1 work, in April 2013 the following properties were added to schema.org to describe items of type Creative Work:

educational alignment: an alignment to an established educational framework. This property can be used to specify that the resource addresses some part of a shared curriculum or competence framework, or that it can be classified against some other educational scheme such as one describing educational levels.

educational use: a text description of the educational purpose of the resource, for example 'assignment' or 'group work'.

interactivity type: the predominant mode of learning supported by the learning resource. Acceptable values are 'active', 'expositive' or 'mixed'.

is based on url: a link to a resource that was used in the creation of the resource that is being described. This may be useful if a learning resource is a derivative of another resource.

learning resource type: a text description of the predominant type or kind characterizing the learning resource. For example, 'presentation', or 'handout'.

time required: the approximate or typical time it takes to work with or through this learning resource for the typical intended target audience.

typical age range: the typical range of ages of the content's intended end user.

Implementation of LRMI metadata
Having extended the schema.org ontology so that the educationally significant properties of a resource can be described, the next challenge is for search services and publishers of information about learning resources to implement LRMI metadata. Phases two and three of the LRMI project are addressing these challenges. Funding has been provided to publishers of Open Educational Resources to implement LRMI/schema.org metadata in their content management and dissemination systems. The Learning Registry, which provides a mechanism for sharing information about learning resources (Learning Registry, n.d.), is exploring the use of JSON-LD for sharing LRMI metadata as stand-alone records rather than embedded in the HTML of web pages so that it may be used in resource discovery by search services that are connected to the Learning Registry. Finally, as the search services most people use to find educational resources are the big search engine providers: Google, Bing, Yahoo, Yandex, Baidu; LRMI has an

advantage over previous attempts at enhancing educational resource discovery, as it is based on a technology that is developed, endorsed and promoted by four of these companies. That is no guarantee that they will see a compelling business case to create tools and services that exploit LRMI metadata for the discovery of learning resources, however their support for the schema.org syntax and ontology means that any information encoded in webpages using LRMI metadata will be available through some of their generic tools and services, e.g. Google custom search engines.

Through the efforts of the LRMI project and those implementing LRMI metadata, we are seeing large numbers of resources being described with open machine readable metadata, which relates to the criteria used when learners and educators search for open educational resources. It is our hope that continued uptake of LRMI metadata can be further stimulated by increased awareness among those who commission and manage resources for educational use and who are in a position to request that the resources and services created for them are described with LRMI metadata. A virtuous cycle of expectation and creation of well described resources would help ease some of the problems that currently exist around the discovery of resources that meet specific individual learning requirements.

References
Learning Registry (n.d.) Learning Registry website [online]. Available from http://learningregistry.org/ Accessed: 29 November 2013
LRMI (n.d.) About the LRMI [online]. Available from http://www.lrmi.net/about Accessed: 29 November 2013
schema.org (n.d.) What is Schema.org? [online]. Available from http://schema.org/ Accessed: 29 November 2013