I2G

INTERGEO

Deliverable Nº: D2.4

Metadata Specification

The INTERGEO Consortium

June 2008

Version: resubmitted version Dec 5th 2008

Main Authors:

Maxim Hendriks (Technische Universiteit Eindhoven)
Paul Libbrecht (DFKI GmbH)
Albert Creus-Mir (Maths 4 More Inc.)
Michael Dietrich (DFKI GmbH)

Project co-funded by the European Community under the eContentplus Programme

Project ref.no.	ECP-2006-EDU-410016
Project title	INTERGEO - Interoperable Interactive Geometry for Europe

Deliverable status	submitted
Contractual date of delivery	M9 – June 2008
Actual date of delivery	June 30^{th} 2008
Deliverable title	Final Metadata Specification
Type	Report
Status & version	resubmitted version Dec 5^{th} 2008
Number of pages	36
WP contributing to the deliverable	WP2
WP/Task responsible	Paul Libbrecht
Authors	Albert Creus-Mir (Maths 4 More Inc.), Michael Dietrich (DFKI GmbH), Maxim Hendriks (Technische Universiteit Eindhoven), Paul Libbrecht (DFKI GmbH)
EC Project Officer	Spyridon Pilos
Keywords	metadata, interoperability, LOM, user-input, search
License	This document is available under the license *Creative Commons Attributions Sharealike Germany 2.5* [Cre08]

Contents

1 Executive Summary

This report describes the metadata schema which the INTERGEO project will be using as basis for the annotations of learning resources in its platform dedicated to the sharing of interactive geometry constructions and related resources.

The metadata schema is an application profile of the widespread Learning Objects Metadata standard of IEEE [Com02] which specifies a set of fields, its meaning for the general purpose of tagging learning resources, as well as an XML specification to encode it. Our specification extends the LOM specification and its XML schema binding [Com] which provides approaches towards interoperability with learning management systems and with other learning objects' publications on the web.

The INTERGEO specification is both smaller than IEEE LOM and extends it, as most application profiles do: the genericity of LOM makes it too hard to be used by common-users in its entirety, moreover, special functions often require addition metadata fields. The simplification go in the direction of the specialization for interactive geometry constructions; the extensions go in the direction of either the interactive geometry specialization or the fine-grained annotations required by the objective of cross-curriculum search where, mostly, the GeoSkills ontology plays a role [Alb08].

This deliverable is organized as follows: after an introduction, we survey the related metadata specifications that apply to our project, then, the objectives and usage of the metadata structures are explained which gives a good hint about the potential of this specification. The individual fields are then specified with their possible values. Finally the measures taken to be taken in order to raise the quality of metadata records are described.

This metadata specification enters the development process of the work-package 4 of INTERGEO where it will be used to create interactive forms for the metadata edition process, where it will be used for the search engine's population and queries, and in the many other usages described in section 3.

2 Introduction

Metadata defines itself as data about the data... a somewhat abstract concept! Within a platform meant to manage resources, metadata is generally understood as the information that is manipulated outside of the resource itself. Hence, metadata is inheritently bound to the platform and tools that manipulates it. This has given the rise to metadata *application profiles* [HP08] which are documented specializations and extensions of a metadata standard.

The Intergeo platform shall provide users with means to share, catalog, and search educational resources based on interactive geometry, with a particular focus on individual interactive geometry constructions. To this end, users will provide information *about* the resources and this information constitutes, precisely, the metadata.

The objective of a metadata specification is to describe the abstract fields that represent the information that users shall browse, catalog, and query the resources with. This specification is a data-model which extends and specializes a widespread model, the Learning Objects Metadata [Com02]. This extension and specialization follows the approaches of an *application profile* for LOM [HP08] and thus should be able to, partially, interoperate with external systems.

The Intergeo platform shall be based on the Curriki platform, itself an application of the XWiki platform. [1] The Curriki platform is a rich Java Servlet based community server, whose main instance has grown mature with a membership reaching 40 000 users. The Curriki platform has a metadata scheme which is very general in scope, with a very shallow topic organization, no competency annotation possible, and no support for multilingual resources (which most of the interactive geometry resources are). An adaptation of the Curriki platform for the metadata schema described in this report is part of the work of the work-package 4 of the Intergeo project.

The Intergeo search engine has to offer the services of a **cross-curriculum search**: as explained in [LDM+08], this should, for example, allow a Scottish teacher to find a resource contributed by a French colleague. Thus the Intergeo metadata specification needs to include tags that cross the curriculum boundaries: the work done in [Alb08] has identified that competencies, topics, and

[1] More information on Curriki from `http://www.curriki.org/`

educational levels, all of which are metadata slots, will be encoded as elements of an ontology.

This specification is, thus, a specialization of classical metadata to the interactive geometry domain as well as an enrichment for cross-curriculum search.

The Intergeo metadata is made of:

- **bibliographic** information such as authorship, rights, and titles
- **pedagogic simple** information such as interactivity type... their possible value is (mostly) a subset of the LOM possible values
- **pedagogic ontological** information denote fine grained annotations about topics, competencies trained, and intended educational programme: these annotations are identifiers in the GeoSkills ontology [Alb08]

This report is organized as follows: after a short survey of some existing metadata approaches, we present the expected usage of the metadata, under the interoperability and user-interactions perspectives, we then present the fields of LOM that we have identified as relevant and indicate their specialization where needed.

2.1 Existing Metadata Standards

In this subsection, we review some existing metadata standards. It was clear from the outset that we might not find a metadata standard that contained precisely those data elements that we need with the values we want. So we knew we had to look around for a widely accepted standard that would approximate our wishes, and then create an *application profile* of this standard. That standard B is an application profile of standard A means that the values of (part of) the data elements of B can be mapped to schema A. This enables the possibility for people or sites working with standard A to still extract metadata from sites working with standard B. It therefore achieves a measure of interoperability, while enabling the creators of B (our consortium in this case) to tailor the metadata to their specific needs. See also [HP08] and [Eri02].

The first standard to look at was the Dublin Core standard, possibly the most well-known standard in the field of metadata.

2.1.1 Dublin Core

The **Dublin Core** metadata set is a standard used to provide cross-domain information for digital resources like videos, sounds, images, texts or composite medias e.g. a web page. It dates back to 1995 where, in Dublin, Ohio, U.S., an invitational workshop was held by the Online Computer Library Center

(OCLC, `http://www.oclc.org/us/en/default.htm`) in which the work on the metadata standard set was started. Dublin Core provides a basic (or core) set of metadata which can be extended and refined if need be. There are currently two levels of the standard: **Simple Dublin Core** and **Qualified Dublin Core**.

Simple Dublin Core includes the following 15 metadata elements

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Every element in this list is optional and may be repeated as often as needed.

In addition to the above mentioned metadata elements Qualified Dublin Core includes three more metadata elements

- Audience
- Provenance
- RightsHolder

and a group of so-called qualifiers which can be used to refine the semantics of the elements in ways that are useful in resource relocation and discovery [Com00].

As can be seen from the lists above the Dublin Core metadata provides the **bibliographic** information for the given resources. A full overview on the Dublin Core elements and element refinements can be seen at

`http://dublincore.org/documents/dcmi-terms/`

and an RDF/XML encoding at

`http://dublincore.org/2008/01/14/dcelements.rdf`.

Extensibility, or, differently said, adaptability and the specification of adaptations, has only been addressed recently in Dublin Core with the notion of application profile emerging as current working drafts.

At the moment, the *Dublin Core Education Community* is designing a metadata schema which would be an application profile of the Dublin Core Standards specifically geared towards the field of education. However, it hasn't been presented yet, we cannot use it. Luckily, this initiave builds on another standard for metadata that we now describe, that *is* directed specifically towards education.

2.1.2 LOM: IMS and IEEE Learning Objects Metadata

The Learning Objects Metadata standard [Com02], or LOM for short, is a widespread standard the aim of which is to provide a general framework to designate the fields of metadata of learning resources and to encode them. Many fields of LOM extend those of the Dublin Core initiative. That is, they convey information that could be classified with DC elements, but these data elements are defined to contain more specific information. In more technical words, the LOM metadata schema is an application profile of Dublin Core. There are also various data elements in LOM that cannot be seen as specifications of Dublin Core, but give additional information. With LOM, educators can add metadata using standardised values that are clear in meaning to a relatively large community.

The LOM specification comes with an XML schema and has been the subject of several other encoding attempts, in particular in RDF [Com]. The LOM specification is equivalent to the IMS Learning Resource Object Metadata [IMS06].

Thus, choosing LOM as our metadata standard would entail compatibility with Dublin Core and with many from the field of education. LOM has been used by the educational ministries of several countries to help add metadata to computer-based resources. To this end, they have made their own LOM application profiles. See for example LOM-FR (`http://www.lom-fr.fr/`) in France and LOM-ES [ea08] in Spain.

2.1.3 Other standards

Several other metadata standards were briefly looked at, but not found satisfactory. The MODS (`http://www.loc.gov/standards/mods/`) and METS

(`http://www.loc.gov/standards/mets/`) standards were much too general to be of any use. A new standard called MLR (Metadata for Learning Resources) is being developed under auspicions of the International Standards Organization as standard ISO/IEC JTC1 SC36. This would be a competitor to LOM, but has not gained the same acceptance yet. Choosing this standard would mean furthering a splitting of the educational community in different factions. Moreover, in 2008 there seems to be a lot of renewed activity on educational metadata, and the problems and disadvantages of LOM v1.0 that people have noted will be addressed in a new version of LOM. The MLR initiative has now apparently agreed to work together with the IEEE Working Group on a new LOM. More information can be found in [Cur08].

As a result, we chose to base our metadata specification on LOM v1.0. This will probably maximize the amount of metadata we could export to other platforms for interoperability, while giving us the detailed information we want — as we now see it. If a future version of LOM would urge us to alter our standard slightly, we will keep an open mind. We will report on the functioning and possible updating of the standard in November 2009 (Deliverable 2.6).

One practical remark we add about LOM is that it is often considered a somewhat large set of fields. Indeed, full LOM editors such as the RELOAD editor require daunting work from the user. To counter this problem, a number of fields from the INTERGEO application profile of LOM will be filled automatically. The LOM fields are all listed in the section 4 and the fields we have decided to adapt or suppress are indicated.

2.2 Other projects

We also looked at and contacted a few other projects that work towards goals similar to ours. To have metadata interoperability from our system towards theirs would be a nice feature we kept in mind.

2.2.1 Curriki

The Curriki platform (on which our platform software is based) uses a metadata schema that is very broad and thematically shallow. A screenshot of the information panel is presented in figure 1 on page 15. The shallowness of these information panels implies that search is mostly done outside metadata, on text.

2.2.2 MELT

The MELT (**Metadata Ecology for Learning and Teaching**) project wants to enable teachers and pupils in schools finding resources fitting their needs, language, culture and preferred way of teaching and learning. It does not specialize in one particular field of education, but wants to be usable for all fields.

To achieve this goal MELT focussed on creating useful metadata types. This resulted in the MELT's application profile (see [Mir07]).

MELT's application profile is similar to the INTERGEO project metadata, but differs significantly in the way they classify and search resources. In the INTERGEO project resources are classified and later found using an ontology of topics, competencies, and educational programmes (which are, thus, three essential ingredients) whereas MELT seems to be based on the thesaurus classification and still offers tagging as an (extra) option.

The intents of the MELT project seem to overlap strongly with the ones of INTERGEO and exchanges towards a collaboration have been initiated.

3 Metadata Usage

In this section, we present where we expect the metadata records to be used in and how.

3.1 Where and How will Metadata be Found

The prime place for an intergeo metadata record to be living will be within the platform itself where it shall be presented, edited, and queried. The storage will be abstract, based on an efficient data storage and with dedicated presentation and editing forms. Each resource of the platform shall have a unique URL on which is the main link to exchange to denote the resource, as stored in the platform.

Metadata records should, however, be found in three other places:

Web-accessible : aside of being user-browsable and editable on the platform, the metadata shall be also accessible to robots such as OAI harvesters following the OAI-PMH protocol [LdSNW02] which receive metadata records encoded in the INTERGEO metadata format specified below, inherited from the LOM XML schema defined in [Com], or semantic-web crawlers that can read metadata records embedded in web-pages using such approaches as GRDDL [Con07].

Embedded in i2g Constructions : Interactive Geometry tools owe their success, in part, to their easy authoring nature: compared to hand-programmed interactive applets, interactive geometry applets are composed using a geometric tool: the interactive geometry system itself and its ability to publish files for "players" to execute them (plugins, Java Applets, Flash player...). The i2g format, delivered by the work-package 3 of Intergeo will provide space in the format for the metadata to be embedded, in its XML or RDF format.

Thus the metadata will be editable within the dynamic geometry system, just as it is presently the case for several dynamic geometry systems in which the author of a construction can edit some metadata fields within text fields and save them while saving the file. The metadata will, then, also be publishable within the i2g format.

Linked to Web-Pages : Interactive geometry constructions will not only be deployed on the Intergeo platform and, most probably, the web shall display collections of interactive geometry constructions which are not (yet) indexed within the Intergeo platform.

In order for these resources to be easy to re-use, they should be linked to their metadata record. The embedding described above, i2g-metadata (preferred) and GRDDL, are all applicable and will be put to use so as to ease up the introduction into the platform of a resource already published elsewhere on the web.

3.2 Interoperability Potential

Because the metadata is expressed in standardized encodings, it may be used in other tools.

Harvesters are on such a sort of tools, be them RDF, OAI, or Intergeo oriented: they generally offer search services and, as such, it is important for them to provide all possible metadata. It is not clear yet whether a rich educational model will benefit of this interoperability or whether only the bibliographic model of Dublin-Core will be exposed.
An implementation of the OAI-MHP protocol is planned in the Curriki platform which will include the Dublin Core and LOM flavours.

SCORM Packagers and Learning Management Systems : LMSs are systems that connect organizational and communication aspects of the life of an educational institution with learning contents: they mostly present content embedded within SCORM packages [DT06] which are assembled by dedicated authoring tools.

The metadata records, in XML format, shall be embedded in the SCORM packages exported from the intergeo platform. Both of these features are in the roadmap of the Curriki platform.

In this interoperability case, it is expected that the bibliographical metadata and pedagogical metadata may be of interest (for example a SCORM player may simply use the icons to represent the types of resources), but that the information of ontological nature is unlikely to be compatible.

Semantically Interoperable Repositories : we anticipate the rise of systems that are able to read the metadata information not only in their bibliographical and pedagogical aspects but also in their ontological aspects: the usage of OWL reasoning or of OWL translation tools provide a very strong potential for translations to other formalisms. Such a vision is described in the *ecological approach* of G. McCalla [McC04] for example.

Together with the OFSET group,[2] we intend to experiment such an interoperability by letting the PHP-based GNU edu repository platform[3] interoperate with the whole of the metadata records in particular the competency and topic records both of which have a matching metadata slot in GNU edu: skills and keywords.

3.3 Interactions with Metadata

Metadata shall be processed by several users using the following user-interface artifacts.

Metadata View The information stored in the metadata record shall be rendered in order to provide information about a resource, this is the metadata view. This rendering will be accessible in a panel named "Information" in the resource web-page. All the ontological information in the record shall be connected to ontology navigators. An example view as done currently on `http://curriki.org` is displayed in figure 1.

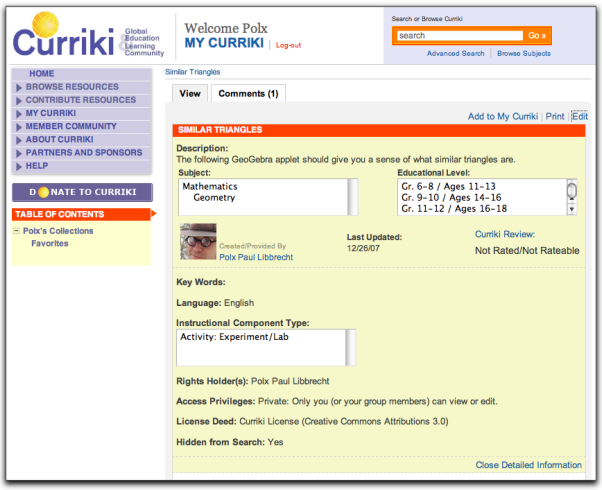

Figure 1: Curriki Metadata view.

Metadata Editing When contributing a resource, an editing panel will be presented to the *annotator* user. This editing facility will be mostly made of

[2]The OFSET group is a non-profit organization of teachers for the Open-Source development of IT in education. More information about them is accessible from `http://www.ofset.org`.

[3]GNU edu is at `http://gnuedu.ofset.org/`.

text-fields with links to explanations, for slots which have values in a small set, a pop-up menu shall be offered, for slots which have values in a large set (mostly the ontological slots), an autocompletion and navigation picker will be offered.

This editing form has to support validation of all fields in order to let the user ensure the quality of the metadata record. Validation is to be done at input time as well as using batch validating tools which are run by editors that review the content of the database, see section 5.

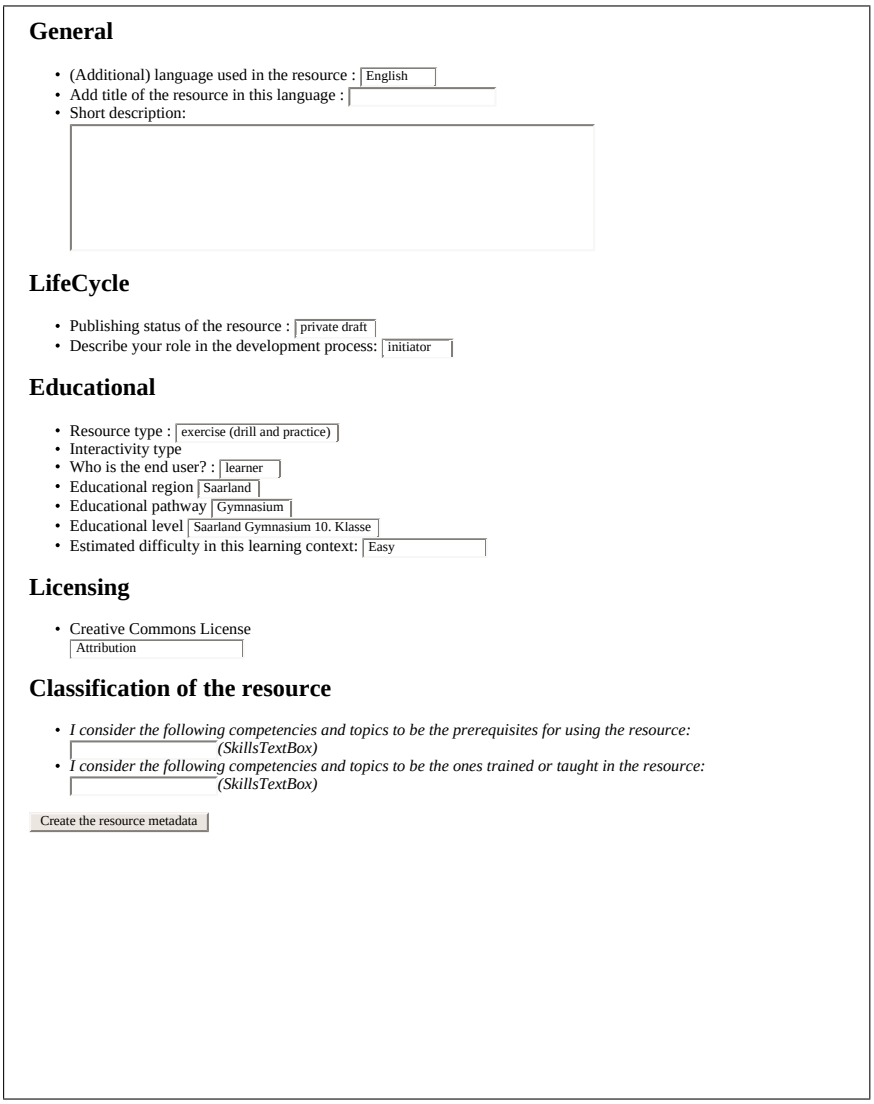

Figure 2: A prototype of the metadata form that authors will be using. We aim for simplicity.

Query Forms Searching through the metadata shall be done with the simple text field classically used in search engines.

Searching shall be helped by an autocompletion mechanism described in [LDM+08] which will suggest possible completions for words that are partially entered among the labels of the topics, competencies, and educational levels as well as the names of authors.

Using this autocompletion allows a precise designation of the entity indicated and thus avoids the need to learn a more elaborate search with typed queries where the metadata field being searched is made explicit. This latter form will still be made available using the classical "key:value" syntax, as in "author:Nicolas Bourbaki".[4]

Several queries shall also be offered by the way of visual navigation: the ontology navigation display at a node shall link to the search for the resources relevant to this node and shall offer to insert itself as an extra search criterion.

The search engine of Intergeo is described in more detail in [LDM+08] and in [Alb08]. It will be based on classical information retrieval principles and tools which provide infrastructure for large scale ' indexing and queries supporting:

- tolerance to typos (matching with lower score a word that is mistyped) thanks to the edit-distance matching algorithm of Lucene
- stemming (inducing tolerance to plural/singular and verb-termination differences), thanks to the Snowball family of analyzers for Lucene
- input problems with internationalized characters (matching, with lower score, `gmuend` for gmünd, `ete` for été, and `espanha` for españa, by following strategies similar to those in the Accents Drupal module.
- phonetic matching (matching the words that *sound the same*) with the Phonetix metaphone analyzer

The search tool is under development in work package 4 of INTERGEO.

[4] A generic example of `key:value` query syntax is documented in Apache Lucene's query syntax page: `http://lucene.apache.org/java/2_3_2/queryparsersyntax.html`.

4 Metadata Fields Specification

This section details the LOM specification and how we are planning to use and subtly adapt this to our needs. The LOM standard defines nine classes of data elements. These are the following:

Group of LOM data elements	**Main content (within the I2G context)**	**Edited by**
1 General	Resource bibliographic data	hand
2 Life Cycle	History of the resource	hand
3 Meta-Metadata	History of this metadata	auto
4 Technical	Resource file and system compatibility properties	auto
5 Educational	Info on the educational content	hand
6 Rights	A Creative Common License	fixed
7 Relation	-	ignored
8 Annotation	Comments by users	auto
9 Classification	Links to ontology	hand

Figure 3: Summary of our LOM sections

All data elements are presented and explained when they are not fully copied from LOM as can be browsed at [Com02]. We could not, unfortunately, provide fine-grained links from our specification to the LOM specification. A web-published version of this document is fine-referenceable so that developers and

tool designers can refer to it: for example, the anchor name of the field 1.1:Identifier is #`1.1:Identifier`.

We now discuss all the data elements in each of these nine groups.

4.1 LOM.General

1.1 Identifier

This will be a short list of labels, each uniquely identifying the resource. We will use at least two instances of this field. One will be an automatically generated numerical entry. The other will be a more conceptual human readable entry, namely a URI of a form similar to `http://i2geo.net/resources/Thales_theorem`, such as the URI of a wiki page. This entry will be generated based on the title of the resource. So 1.1.1-Catalog will consist of "`http://i2geo.net/`" and 1.1.2 of will be of the form "`http://i2geo.net/resources/my_theorem`".

1.2 Title

This will contain a list of LangStrings with the title in various languages. Users will be able to add a title in their own language.

Example

```
<title>
    <string xml:lang="en">Interception theorem</string>
</title>
<title>
    <string xml:lang="fr">Thï¿½orï¿½me de Thalï¿½s</string>
</title>
```

1.3 Language

A list of the languages supported by the resource. The value "none" is also allowed to mean there is no language used. This field being left out will mean that the language used is in some sense universal, that everybody (in the EU at least) will be able to work with it.

1.4 Description

This field will contain concise descriptions of the resource, in various languages, each a LangString. There will be at most one for each language.

1.5 Keyword

This field is not of use to the INTERGEO platform itself, since we have implemented an ontology to characterize content of resources (see 9:Classification and

Deliverable 2.3 [Alb08]). However, for the sake of interoperability with other platforms that might want to use the INTERGEO resources, a set of keywords will be generated from the ontological annotation of the resource.

1.6 **Coverage**
left out

1.7 **Structure**
left out

1.8 **Aggregation Level**
left out

4.2 LOM.LifeCycle

2.1 **Version**
auto-generated
The version number of the resource. We will adhere to numbering with natural numbers, starting at 1. This is not as general as LOM allows for this data element, but makes for maximal clarity.

New (2.2*) **PublishingStatus**
user-private draft | *group-private draft* | *public draft* | *published* | *retracted*

We want authors to have control over the visibility of the resources they contribute. This can be handled with the PublishingStatus data element. Users can make the resource a user-private draft, which will only be visible by themselves. Moreover, we will have the possibility for users to create groups for more efficient collaboration and exchange. Thus we will allow the possibility of group-private drafts, visible only within such a group. Public drafts will be visible to everyone, but with indication that they have not been finalized.

2.3 **Contribute**

This field (its subfields) collect references to the people that contributed to the resource:

2.3.1 **Role**
author | *editor* | *annotator* | *validator* | *retractor*
The author will be the original designer of the resource. An annotator will be someone writing descriptions or comments, a validator someone proofing and adjusting the metadata record and a retractor somebody who (temporarily) took this resource out of the publicly visible database.

2.3.2 **Entity**

This will define who the contributor is. We will let contributors fill in a form, allowing them to enter only very little personal data, but still making it possible to point to a unique place where (some) data on this person can be found. This data will then be stored in the form of a vCard. The contributor field will then contain the URI of this vCard, e.g. `http://www.i2geo.net/users/myself`.

2.3.3 **Date**
auto-generated
The date of the initiation/edit/annotation/retraction of the resource. The system will default to "now", but the contributor has the final say.

4.3 LOM.MetaMetadata

In this section, the metadata record itself is classified, not the resource. This is the reason there are also identifiers and contributors in this section.

3.1 **Identifier**
auto-generated
A unique identifier for this metadata record itself, just like 1.1 is for the resource.

3.2 **Contribute**
auto-generated
Analogous to 2.3, this field and its subfields will contain data in the form of a vCard of the contributor, but in this case of a contributor for this metadata itself. We note that the roles we allow here are "initiator" and "editor".

3.3 **Metadata Schema**
auto-generated
This will contain the URI of the RDF/XML specification of this metadata standard.

3.4 **Language**
auto-generated
This states the default language of this metadata record itself. We will prefill "en" (English) here. A french description in 1.4:Description, for example, would be added as a LangString ("fr","La plus belle resource du monde") so the language would be clear.

4.4 LOM.Technical

4.1 **Format**
auto-generated
The name of the format, which we will supply for standardization in Work Package 3 or the mime-type of the document.

4.2 **Size**
auto-generated

4.3 **Location**
auto-generated
The URI to the place on the INTERGEO platform where this resource can be found. Will be automatically computed, based, in our case, to the Identifier 1.1.

4.4 **Requirement**
auto-generated
This is a disjunction of all the technical requirements for the use of this resource. Since operating system and browser, which LOM gives as default value space, do not give compatibility issues for our purposes, we define our own value space here with type "DGS" and tokens the names of various DGS. This must be seen as a key that can have as value a set of names of the Dynamic Geometry System with which the resource is supposed to work, including the minimally required versions.

4.5 **Installation Remarks**
auto-generated
We will fill this with a standard text similar to: "Load the resource into a Dynamic Geometry System supporting the INTERGEO format. For a list of these programs, other resources, and help, see http://i2geo.net/".

4.6 **Other Platform Requirements**

Users may put extra remarks here about the functionality of this resource. This may concern functionality within different DGS, e.g. that this specific resource will only work in a limited way with a specific DGS, or it may concern other unforeseen future requirements.

4.7 **Duration**
left out
This element is not applicable as such to dynamic geometry constructions.

4.5 LOM.Educational

5.1 **Interactivity Type**
active | *expositive* | *mixed*
This is the mode of learning the resource supports. This has much to do with 5.3:InteractivityLevel. Most constructions will have active type, while some other documents on the platform, like curriculum standards, might get expositive.

5.2 **Learning Resource Type**
exposition | *drill_and_practice* | *investigation* | *assessment* | *other*
The standard value space that LOM offered seemed overly complicated to us, and not to the point. Based on the experience of Cabrilog and OFSET, we defined the simple value you see here. It mainly speaks for itself. The difference between drill_and_practice and exploration is that resources of the latter type will are more open-ended.

5.3 **Interactivity Level**
very low | *low* | *medium* | *high* | *very high*

5.4 **Semantic Density**
left out

5.5 **Intended End User Role**
learner | *teacher* | *author* | *manager*
Most resources are probably either for learners or teachers. But some other documents on the platform might be for managers.

5.6 **Context**
school | *higher education* | *training* | *other*
LOM suggests putting two instances of this data element, one for coarse categorisation, one for fine-grained. We will use ("LOMv1.0","school") for the coarse one.

For the fine grained one we shall use the URI of the OWL individual representing that educational context, an instance of ◇ `EducationalLevel` and ◇ `EducationalProgram`. This means a possible value could be ◇ `Gymnasium_Saarland_9te` which should be rendered as "Gymnasium Saarland 10. Klasse".

5.7 **Typical Age Range**
auto-generated
We let this range be computed from the ontology and the set of Educational Levels that have been entered for this resource in 5.6:Context, which we view as more detailed information. This field is kept for interoperability.

5.8 **Difficulty**
very easy | *easy* | *medium* | *difficult* | *very difficult*
Difficulty for target audience. We will try to let this sort itself out and let the creator and later validators of a resource assert how difficult a resource is. We expect this field to be useful for teachers: for example, they often want to know whether a resource is suitable for a first aquaintance with the theme of the resource.

5.9 **Typical Learning Time**
left out
5.8:Difficulty says enough already.

5.10 **Description**

We did not see the use of this data element over 1.4 together with LOM.Annotations (which we will handle differently, as described below).

5.11 **Language**
left out

4.6 LOM.Rights

6.1 **Cost**
no
The resources on the INTERGEO platform will not cost the user anything.

6.2 **Copyright and Other Restrictions**
yes
LOM allows an author to express here whether some form of copyright applies to the resource. To the resources on the intergeo platform, there will *always* apply some form of copyrights, the default being CC-Attribution-Sharealike [Cre08].

6.3 **Description**
auto-generated
Automatically filled after the choice for a specific license. A human-readable description and upshot of this license will be put here in English.

New (6.4) **LicenseURI**

Since some form of copyright always implicitly applies to mostly any product, we want to explicitize this by letting authors mandatorily choose a License for the resources they supply. The URI of the Creative Commons license *deed* will be used. For example the URL `http://creativecommons.org/licenses/by-sa/2.0/de/` designates the license of this document.

4.7 LOM.Relation

The relations between resources will be handled by our search engine. We explicitly do not want these to be hardcoded in the resources themselves, so we will leave this part of LOM out of our metadata.

4.8 LOM.Annotation

This part of LOM is supposed to be used for comments by users on their experience with the resource. On the INTERGEO platform, each resource will have its own discussion forum, just like on a wiki. For size purposes, we will not export all the data of the this forum to the LOM metadata of the resource. Instead, we will ask annotators on the platform to submit a grade if they review and annotate a resource. We will then include a list of these grades in the metadata, together with hyperlinks to the corresponding comments on the platform.

8.1 **Entity**
auto-generated
This will contain the URI to the place on the user forum or the quality review system for the resource where this annotation can be found.

8.2 **Date**
auto-generated

8.3 **Description**
left out

New (8.4) **Quality**
auto-generated
This will contain the short quality assessment of the annotator in the form of a number between 0 (very very bad) and 100 (very very good). For flexibility, we will allow a float as number, see `http://www.w3.org/TR/xmlschema-2/#float`.

New (8.5) **NumEvaluations**
auto-generated
This field contains the number of evaluations used in the quality assessment whose quality is reported. Using Quality and NumEvaluations allows the platform to export an aggregation of many quality results which will be needed as soon as more than 20 evaluations will be done.

4.9 LOM.Classification

A resource will be classified with our ontology. This will be done with the competencies and topics described in [Alb08]. The resulting classifying expression will be put in 9.2.2

9.1 **Purpose**
prerequisite | educational objective
We deviate from the standard LOM value space and restrict this to { prerequisite , educational objective }. This means that a topic or competency from 9.2 is either presupposed to work with this specific resource, or is trained/taught in this resource.

9.2 **Taxon Path**

The users will choose the competencies and topics with our SkillsTextBox utility, described in [Alb08] and [LDM+08]. The subfields will then be filled automatically.

The Id child of the Taxon element inside taxonPath is the URI of the OWL individual representing that topic or competency. Such URIs are provided, as of today, in the OWLdoc view of the ontology on `http://i2geo.net/ontologies/dev`.

For example the competency ⋄ `Calculate_sum_of_fractions` has as canonical URI:
`http://www.inter2geo.eu/2008/ontology/ontology.owl#Calculate_sum_of_fractions`

but would be shown to the users as one of "calcular la suma de dos fracciones", "calculate sums of fractions", "Summen von Brüchen berechnen", or "Addition des fractions".

9.3 **Description**
left out

9.4 **Keyword**
left out

4.10 Added Fields

As we have seen above, two new data elements have been defined for INTERGEO, namely

- (2.2*) PublishingStatus
- (6.4) LicenseURI

Moreover, we have adapted or created the following value spaces to our needs:

- 2.2 PublishingStatus: { private draft, public draft, published, retracted}
- 2.3.1 Role : { author, editor, annotator, validator, retractor}
- 3.2.1 Role: { initiator, editor}
- 5.2 Educational resource type: { exposition, drill_and_practice, investigation, assessment, other}
- 5.5 Intended End User Role: { learner, teacher, author, manager}
- 5.6 Educational context: the URI of an OWL individual representing the intended educational context or educational level
- 8.4 Quality: a float between 0 and 100.
- 8.5 NumEvaluations: an integer representing the number of evaluations behind the given quality statement
- 9.1 Purpose: { prerequisite, educational objective }

5 Strategies for Metadata Quality

In this section we propose methods to alleviate the work of metadata input of users while striving for completeness.

On the one hand, automated validation criteria are offered, which will be checked either at input or analysis time. On the other hand, the quality of a resource displayed on the intergeo platform will be made up by the quality the resource itself as well as the one of its metadata. The quality framework [MSLT08] includes, indeed, feedback and workflows for the evolution of the metadata. Finally, the work of metadata input can be alleviated a lot thanks to suggested values *prefilled* by context and content. Below we describe the validation criteria and the values suggestions.

5.1 Validation Criteria for intergeo Metadata

Validation criteria are, on the one hand, on single metadata fields which is documented in the previous section. Single field validation can be:

- enforce the membership of a controlled vocabulary (also called *symbolic*) (which can be generally ensured using user-interface artifacts but may still need error reporting as the controlled vocabulary may be evolving).
  The values taken are generally identifiers to the vocabulary entries (typically a URI), and the user-interface needs to render them in the appropriate language. Controlled vocabularies of LOM include include:
  - the 5.6:Context
  - the Classification/TaxonomyPath expects value as individuals of the ◇ `Competency` or ◇ `Topic` classes of GeoSkills
  - many other fields have their list of possible values documented above.
- character strings criteria which are such as "non-empty", "all letters and...". Such constraints are generally expressed using regular expressions in the XML schema and we simply repeat the ones of LOM as expressed in [Com].

More elaborate validation strategies, validating several fields together, will appear, based on experience of the contributions.

5.2 Review Based Validation

As part of the work-package 6, usage of a resource could be followed by a quality report which may also lead to changes to the resource or to its metadata. Indeed, the questionnaire of [MSLT08] has a block of questions about metadata and we expect it very possible that some of the fields of the metadata will be reviewed by users different than the original metadata contributor. In some cases, however, it will not be possible to change the metadata in a compatible way and it will then be possible to add another metadata record describing the same resource.

Some possibilities of conflicting values given to metadata slots might occur for instance when specifying the interactivity type or interactivity level. Let us give the example of a resource displaying two secant lines cut by parallel lines and ratios of corresponding segments. A displayed text formulates the theorem and mentions "you can drag points a, b, c and lines s and r and observe the equality of ratios". The author probably assigned the value "active" to the interactivity type because the student is requested to move elements of the figure. A reviewer may find that the student is not really active mathematically speaking since s(h)e has just to observe that the equality of ratios is always valid when moving points or lines. This reviewer may consider that the interactivity type is rather expositive. All information about the theorem is given and students have just to check it by dragging.

In the same manner, the field Learning Resource Type that can be filled with exposition/drill_and_practice/ exploration/assessment/other may give rise to conflicting values for the same resource depending on the point of view of the reviewer. Dragging and checking the validity of a specified theorem may be judged as exploring for some reviewers and as exposition by some others.

The INTERGEO platform will allow the reviews on the constructions as described in [MSLT08]; the reviews are both about the resource itself and the composite of the resource and the metadata record. This review will be filled and questioned inside the INTERGEO platform based on the questionnaire available at [MSLT08]. The summarization of the quality reports will flow into section 8 of the metadata and will, thus, influence the search engine's ranking so that resources with a *higher quality rank* will rank higher in the search score and thus may come first on the results list.

5.3 Suggestions Prefills for intergeo Metadata

As fundamental help for the population of the metadata of the learning resources appears to be the service of *metadata generation* which is addressed by several projects including that of GNU edu: its prefilling strategy is not documented yet but we obtained the following strategy which describes well the general purpose: "Upload from a URL: The system tries to download the file. This is a useful check in itself, because any error in file location is harmful. Additionally, the

system can immediately guess the file format and compute the file size. Then it analyzes the headers: ¡meta¿ in normal HTML, \header in WIMS,... This helps to guess title, author, language, keywords, description. For Wikipedia pages, GNU Edu adds such data like license, author, email and learning resource type ; for WIMS interactive pages, it deduces the school level and subject : such strategy can be adapted for other groups of authors.".

Other projects focussed on metadata generation include:

- KU Leuven's Metadata Generation at `http://www.cs.kuleuven.be/~hmdb/joomla/index.php?option=com_content&task=view&id=29&Itemid=57`
- OpenCalais `http://www.opencalais.com/calaisAPI`

and several others in the areas of semantic desktop.

Metadata generation can be a process occurring fully under-cover without user visibility, this runs, however, a high-risk of leaving inaccuracies untouched. In INTERGEO, metadata generation will be offered as a suggestion prefill, where the form remains visible but with fields already filled. The contributor will be able to modify these automatically detected values and tune them. In turn, the metadata generation will be able to take advantage of the user validation (have seen but not change the metadata suggestion) or of the correction (the suggestion was not correct).

Metadata generation has two main sources:

Suggestions Based on Content : given a resource as a file to upload or as an HTTP URL, the platform has the potential of detecting a large amount of variables and suggest them: this includes first the presence of a metadata record embedded in the resource (see section 3.1), the format and the technical requirements, topics inferred from the content (e.g. based on the mathematical constructions and the text), or language (based on textual analysis). Most probably, the set of heuristics offered here will grow with the types of resources and the tools available to analyze them.

Suggestions Based on Context of the Contributor : The context of a contributor includes wide array of variables, made of his/her history and profile, as well as the storage of all resources.

The global storage will provide suggestions for identifiers (ensuring uniqueness) at least and will use platform defaults (e.g. the default license).

The profile of the contributor defines his language of use as well as the educational region, pathway, and programmes he is active in, which can define the language and learning-context for the resource.

Finally, the history of the contributor is probably the best source for a low-evidence suggestion: several values may just be suggested to be the last value input: this is certainly relevant for author name and license but is of low evidence for such suggestion as the interactivity-type or topic.

6 Conclusion: Implementation Status

The content of this report presents a detailed description of the metadata fields and of ways they will be used. This report is the result of careful evaluation of the team of the work-package 2 during several teleconferences and following refinements.

The finalized schema specification above is being implemented as a modification of the Curriki metadata information editors and viewers. Part of this implementation depends on the navigation facilities of GeoSkills which are developed as part of the deviation request sketched in [Alb08].

An integration of the metadata validation facilities for the purposes of a revision following a quality feedback will be embedded as the quality framework evolves in maturity during the second year of life of the INTERGEO project.

The export facilities and, more importantly, the import and metadata suggestions facilities will be progressively incorporated into the INTERGEO contribution facilities. The heuristics that make up these suggestion mechanisms will be inserted after the editing workflow is mature.

7 Acknowledgements

The authors would like to thank Odile Bénassy, of OFSET, for her advise on the LOM standard and the best ways to specialize it towards the needs of INTERGEO, Colette Laborde, of Cabrilog SAS, for her pedagogical and interactive geometry insights, and Cyrille Desmoulins, from the Université Joseph Fourier, for advice on ontologies.

8 Bibliography

References

[Alb08] Albert Creus-Mir, Cyrille Desmoulins, Michael Dietrich, Maxim Hendriks, Colette Laborde and Paul Libbrecht. Deliverable d2.3: Internationalized ontology. Deliverable, Intergeo Project, Available at `http://www.inter2geo.eu/files/D2.3-Intl-Ontology.pdf`, May 2008.

[Com] IEEE Learning Technology Standards Committee. IEEE P1484.12.3, Draft 8: Draft Standard for Learning Technology — Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata. Available at `http://ltsc.ieee.org/wg12/files/IEEE_1484_12_03_d8_submitted.pdf`.

[Com00] DCMI Usage Committee. Dublin Core Qualifiers, 2000. See `http://dublincore.org/documents/2000/07/11/dcmes-qualifiers`.

[Com02] IEEE Learning Technology Standards Committee. IEEE Standard for Learning Object Metadata. Number 1484.12.1-2002. See `http://www.ieeeltsc.org/standards/1484-12-1-2002/`., 2002.

[Con07] Dan Connolly. Gleaning resource descriptions from dialects of languages (GRDDL). W3C Recommendation, World Wide Web Consortium, See `http://www.w3.org/TR/grddl/`, 11 2007.

[Cre08] Creative Commons Inc. (CC). Namensnennung-Weitergabe unter gleichen Bedingungen 2.0 Deutschland. Available on the web `http://creativecommons.org/licenses/by-sa/2.0/de/`, May 2008.

[Cur08] Sarah Currier. Metadata for learning resources: An update on standards activity for 2008. *Ariadne*, 55, April 2008.

[DT06] Philip Dodds and Schawn E. Thropp. SCORM 2004, 3rd ed., 2006. See `http://www.adlnet.gov/`.

[ea08] J.J.L. Villalobos et al. Perfil de aplicación LOM-ES1 v.1.0. Technical Report GT9 / GT8 — SC 36/AENOR, AENOR, Available at `http://www.educaplus.org/documentos/lom-es_v1.pdf`, March 2008.

[Eri02] Erik Duval, Wayne Hodgins, Stuart Sutton and Stuart L. Weibel. Metadata principles and practicalities. *D-Lib Magazine*, 8 (4), April 2002.

[HP08] Rachel Heery and Manjula Patel. *Application profiles: mixing and matching metadata schemas*, volume 25 of *ARIADNE*. UKOLN, Available at `http://www.ariadne.ac.uk/issue25/app-profiles/`, April 2008.

[IMS06] IMS Global Learning Consortium. Learning resource metadata specification. Technical report, Available at `http://www.imsglobal.org/metadata/`, August 2006.

[LDM+08] Paul Libbrecht, Cyrille Desmoulins, Christian Mercat, Colette Laborde, Michael Dietrich, and Maxim Hendriks. Cross-curriculum search for intergeo. In Serge Autexier and Masakazu Suzuki, editors, *To Appear in Proceedings of MKM 2008*. Lecture Notes in Artificial Intelligence, Springer Verlag, July 2008.

[LdSNW02] Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. The open archives initiative protocol for metadata harvesting. Technical report, The Open Archive Initiative, 2002. Available at `http://www.openarchives.org/OAI/openarchivesprotocol.html`.

[McC04] Gordon McCalla. The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learner. *Journal of Interactive Media in Education*, 2004-07, May 2004.

[Mir07] Miriam Olga Cecilia, Agustín Muñoz, Mariano Sanz and Frans van Assche. Melt application profile. Deliverable D5.2 Part 3, MELT Project, Available at `http://info.melt-project.eu/shared/data/melt/MELT_D5P2_Part3_final.pdf`, March 2007.

[MSLT08] Christian Mercat, Sophie Soury-Lavergne, and Jana Trgalova. Deliverable d6.1: Quality assessment. Technical report, Intergeo Project, Available at `http://www.inter2geo.eu/files/D6.1_060508.pdf`, March 2008.

9 Annex I — An XML example

Below we present the XML document of the metadata of a typical resource.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- This example is now validated with a LOM XSD Schema.
 In the future, must be validated with our own schema -->
<lom xmlns="http://ltsc.ieee.org/xsd/LOM"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://ltsc.ieee.org/xsd/LOM lomcustomco.xsd"
    xmlns:i2g="http://i2geo.net/namespace">
  <general>
    <identifier>
      <catalog>"The I2G catalog"</catalog>
      <entry>2008-06-20.73689</entry>
    </identifier>
    <identifier>
      <catalog>"URI"</catalog>
      <entry>http://i2geo.net/resource/entfernung</entry>
    </identifier>
    <title>
      <string language="de">Entfernungen eines Punktes von einer Geraden / Kreislinie</string>
      <string language="en">Lenght between points and lines / circumferences</string>
      <string language="es">Distancias entre puntos y rectas / circumferencias</string>
    </title>
    <language>de</language>
    <description>
      <string language="de">Drei Beispiele zum Thema Entfernungen:
        1) Das Lot als kï¿½rzeste Verbindung eines Punktes mit einer Geraden.
        2) Der Radius als kï¿½rzeste Entfernung des Mittelpunkts zur Kreislinie.
        3) Wann ist ein Punkt auï¿½erhalb bzw. innerhalb eines Kreises?</string>
    </description>
    <keyword><string>distance 2D point line circumference</string></keyword>
  </general>
  <lifeCycle>
    <version><string>2</string></version>
    <status><source>i2g</source><value>published</value></status>
    <contribute>
      <role><value>author</value></role>
      <!-- must be a vCard <entity xsi:type="URI">http://i2geo.net/user/Heiko_Vogel</entity> -->
      <i2g:entityURI>http://i2geo.net/user/Heiko_Vogel</i2g:entityURI>
      <date><dateTime>2008-06-19T15:21:00</dateTime></date>
    </contribute>
    <contribute>
      <role><value>editor</value></role>
      <i2g:entityURI>http://i2geo.net/user/Ivan_Altarriba</i2g:entityURI>
      <date><dateTime>2008-06-20T14:11:00</dateTime></date>
    </contribute>
  </lifeCycle>
```

```
<metaMetadata>
  <identifier>
    <catalog>"The I2G metadata index"</catalog>
    <entry>2008-06-20.44236</entry>
  </identifier>
  <contribute><role><source>i2g</source><value>creator</value></role>
    <i2g:entityURI>http://i2geo.net/user/Ivan_Altarriba</i2g:entityURI>
    <date><dateTime>2008-06-20T14:23:00</dateTime></date>
  </contribute>
  <metadataSchema>"The I2G metadata standard"</metadataSchema>
  <language>en</language>
</metaMetadata>
<technical>
  <format>application/xml+i2g</format>
  <size>92347</size>
  <location>http://i2geo.net/resource/2008-06-20.73689</location>
  <requirement>
    <orComposite>
      <type><source>LOMv1.0</source><value>browser</value></type>
      <name><source>LOMv1.0</source><value>opera</value></name>
      <minimumVersion>7.0.4</minimumVersion>
    </orComposite>
  </requirement>
  <installationRemarks>
    <string>Please, load the resource into a program supporting the i2g format.
    For help and a list of these, see http://i2geo.net/help</string>
  </installationRemarks>
  <otherPlatformRequirements>
    <string>tested with cinderella</string>
  </otherPlatformRequirements>
</technical>
<educational>
  <interactivityType><source>LOMv1.0</source><value>active</value></interactivityType>
  <learningResourceType><source>i2g</source><value>exploration</value></learningResourceType>
  <interactivityLevel><source>LOMv1.0</source><value>medium</value></interactivityLevel>
  <intendedEndUserRole><source>LOMv1.0</source><value>learner</value></intendedEndUserRole>
  <context><source>LOMv1.0</source><value>school</value></context>
  <context><source>i2g</source>
    <value>http://www.inter2geo.eu/2008/ontology/ontology.owl#Gymnasium_Saarland_9te</value></context>
  <typicalAgeRange><string>10-12</string></typicalAgeRange>
  <difficulty><source>LOMv1.0</source><value>easy</value></difficulty>
  <description>
    <string language="en">Useful to understand the definitions</string>
  </description>
</educational>
<rights>
  <cost><source>LOMv1.0</source><value>no</value></cost>
  <copyrightAndOtherRestrictions><source>LOMv1.0</source><value>yes</value></copyrightAndOtherRestrictions
  <description>
    <string language="en">This resource falls under the Creative Commons CC-BY-SA License</string>
```

```
    </description>
    <i2g:LicenseURI>http://creativecommons.org/licenses/by-sa/2.0/de/</i2g:LicenseURI>
  </rights>
  <annotation>
    <i2g:entityURI>http://i2geo.net/user/John_Smith</i2g:entityURI>
    <description>
      <string language="en">You can move the label to get a clearer picture.</string>
    </description>
  </annotation>
  <annotation>
    <i2g:entityURI>http://i2geo.net/user/Jos%C3%A9+Garc%C3%ADa</i2g:entityURI>
    <description>
      <string language="es">Demasiado fï¿½cil para bachillerato</string>
    </description>
  </annotation>
  <classification>
    <purpose><source>LOMv1.0</source><value>educational objective</value></purpose>
    <taxonPath>
      <source><string>http://www.inter2geo.eu/2008/ontology/ontology.owl</string></source>
      <taxon>
        <id>http://www.inter2geo.eu/2008/ontology/ontology.owl#distance_point_to_line</id>
        <entry>
          <string language="es">distancia de un punto a una recta</string>
          <string language="en">distance between a point and a line</string>
        </entry>
      </taxon>
      <taxon>
        <id>http://www.inter2geo.eu/2008/ontology/ontology.owl#distance_point_to_circle</id>
        <entry>
          <string language="es">distancia de un punto a una circumferencia</string>
          <string language="en">distance between a point and a circle</string>
        </entry>
      </taxon>
    </taxonPath>
  </classification>
</lom>
```