An approach to the quality and reusability of metadata specifications for e-learning objects

Jacek Plodzien and Ewa Stemposz

Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, and

Alina Stasiecka

Polish-Japanese Institute of Information Technology, Warsaw, Poland

Abstract

Purpose – The purpose of this paper is to discuss research that leads to identifying new metadata for learning objects and extending the SCORM standard.

Design/methodology/approach – The research involves a questionnaire to collect data from users of e-learning resources and a statistical analysis of that data. The discussion employs concepts from such areas as e-learning, didactics, and statistics.

Findings – The main finding is the identification of several learning objects metadata, including those that are not defined in current e-learning standards. Some of the new metadata could be introduced to the existing standard metadata categories; the others could be used to form completely new categories.

Research limitations/implications – The general solution has been developed but more work is still necessary.

Practical implications – The ideas discussed in the paper, especially the identified metadata, could be used to extend the standard metadata (e.g. in the SCORM standard).

Originality/value – The paper presents both a method for identifying learning objects metadata and conclusions that, after further development, could offer additional support for users who need to choose a suitable resource.

Keywords Data handling, E-learning, Data analysis

Paper type Research paper

1. Introduction

In recent years we have witnessed a huge growth in the popularity of e-learning. Apart from undeniable advantages, this learning paradigm also has shortcomings; one of the biggest being the lack of direct interaction between the teacher and student. Therefore, the issue of resources, which practically play the role of the teacher in this kind of learning, is very important. They have to be constructed carefully, with a focus on the proper didactic construction to support correct learning processes. This observation leads to the hypothesis that the quality of resources has a key influence on the quality of the entire e-learning process. The hypothesis in turn implicates the following question: how can a person intending to use a given learning resource determine its quality? Many learning resources created nowadays conform to some of the existing
standards: SCORM (SCORM, 2004), LOM (LOM, 2002), and IMS (IMS, 2004). We use SCORM as a framework for our research. In this standard, metadata describing resources are subdivided into categories. Unfortunately, there is no category explicitly intended for storing data concerning resource quality. Some quality information can be deduced from currently existing metadata, but it seems to be insufficient. In our opinion, the prospective user of a given learning resource should have direct access to the information concerning its quality. For that reason, we propose changes to the SCORM metadata definition. We suggest both extending some of the existing categories (e.g. educational) and – based on our experience in the software engineering field – introducing two completely new categories:

1. Quality, which should contain measures for evaluating a resource’s quality. To some extent, the set of measures would be based on the existing SCORM metadata, but also some new aspects should be addressed; for instance, the characteristics of a resource’s didactic structure with regard to the support for correct learning processes.

2. Reusability, which should deal with the reuse potential of a given resource; where the reuse potential means the possibility to use an e-learning resource to create another resource.

In our opinion, both the new categories and modifications to the existing ones would be very helpful for users who need to decide which learning resource should be selected from several resources on the same/similar subject.

The paper is organized as follows. In Section 2, we present our research concerning the issue of the identification of measures to evaluate the quality of e-learning resources. In Section 3, we present important ideas concerning the new extended set of potential measures and the new questionnaire version, which has been developed after analyzing the results of its first version. In Section 4, we discuss one of the most interesting conclusions from our current research, that is, a proposal to extend some of the existing SCORM metadata categories and to introduce two completely new ones. The last section, Section 5, concludes the paper and outlines our future work.

2. Identification of measures to evaluate the quality of e-learning resources

Many authors of e-learning resources assume that in order to construct a resource of good quality it suffices to follow standards that currently exist. According to us, this assumption does not work in practice. We agree with the opinion of didactic specialists, that for learning to be effective and learning processes to be stimulated properly, the structure of a resource should be based on the model of effective learning. The idea of the model is to give equal importance both to the process of information delivering (lecture-based learning) and to the process of knowledge creating (problem-based learning) (Allesi and Trollip, 2001). The model also requires the appropriate sequence of learning processes, which should be: new knowledge presentation, support in knowledge gaining, practice and evaluation. Support in knowledge gaining means that in e-learning, the student is more responsible for his/her results than in the traditional face-to-face teaching/learning approach; hence the support for the student in knowledge gaining is very important (for example, the appropriate structure of resources is a kind of support). In turn, in the evaluation step, two main problems can be distinguished:
(1) The student is supposed to test the gained knowledge on his/her own – in the e-learning approach this step is especially important because the student has no access to the teacher (or such access is limited).

(2) Evaluation should also concern the learning process, to indicate any problems that the student may encounter in his/her self-learning.

To be able to measure the degree of conformance of a resource structure to the model of effective learning, we propose to map each of the above processes onto a sequence of resource components (Stasiecka et al., 2003, 2004, 2006a). For instance, the process of new knowledge presentation implies the presence of an Introduction part that consists of, among other things, components dealing with stimulating and directing as well as those responsible for motivating. Motivating concerns, for example:

- the credibility of new knowledge and improving skills;
- directing the student’s attention towards concepts that are necessary to understand the problem; and
- taking into account cognition procedures that are necessary to build a new knowledge.

Furthermore, Introduction should also include components that specify the complexity of the problem, present interesting positive and negative examples, increase the student’s curiosity, and so forth. As we can see, it is not easy to create correct content, even for such a simple part of a resource as Introduction.

Taking into consideration all the above mentioned problems, we propose to distinguish two levels of e-learning resource components:

(1) level I: Introduction, Main content, Summary, and Evaluation; and
(2) level II: sub-components of the level I components.

Below we enumerate all the components; we have labeled them with unique identification numbers that we will use later in the paper:

(1) Introduction:
   1.1. Abstract and indication of key elements.
   1.2. Focusing on the content.
   1.3. Motivating the student to start using the resource.
   1.4. Definition of didactic objectives.

(2) Main content:
   2.1. Base knowledge.
   2.2. New knowledge.
   2.3. Examples of applying new knowledge in practice.

(3) Summary:
   3.1. Recapitulation.
   3.2. Indicating opportunities for skills and knowledge transfer to a new context.
3.3. Dictionary of key concepts.
3.4. Literature.

(4) Evaluation:
  4.1. Self-evaluation.
  4.2. Problem questions.
  4.3. Feedback.

Below we give some information about those components whose purpose may not be obvious from their names:

1.2. Focusing on the content: presentation of the substantial ideas of the new-gained knowledge.
1.3. Motivating the student to start using the resource: concepts necessary to understand the problems that the resource is intended for; the same concerns interesting examples and indications of how the gained knowledge could support the potential user in his/her career.

2.1. Base knowledge: prerequisites for the resource.
2.2. New knowledge: the core part of the resource; it concerns the knowledge that the resource is intended for.

4.1. Self-evaluation: various kinds of tests supporting the process of the user’s self-evaluation, for instance:
  - Simulation: case studies, role playing, games, guided analysis, etc.
  - Drill and practice: one-choice questions, multiple-choice questions, matching, jigsaw puzzles, open questions, etc.

4.2. Problem questions: problem questions for testing the new knowledge, e.g. solutions to the problems in a new context, evaluating other persons’ solutions, rationale for the selected solution.

4.3. Feedback: indications of how the student can contact the teacher (e.g. chat, e-mail).

For further consideration, we assume correspondence between some features of components (they are component presence in a resource and component quality assessment) and measures that we propose to use for evaluating quality of an e-learning resource (respectively, presence measure and quality measure). Later in the paper, we use the name mark for the quality measure. We consider both the presence of level I and II components in a resource and the assessment of the quality of each component. The values for the measures are taken from the sets:

- \(\{0, 1\}\) for the presence measure (0 for absence, 1 for presence).
- \(\{0, 1, 2, 3, 4, 5\}\) for the quality measure.

Because there are 18 level I and II components, there are \(2 \times 18 = 36\) such measures.

For consistency, following the above assumption, the identification number and the name of each measure are the same as the identification number and the name of the corresponding component. Additionally, for measures we use suffixes: the \(p\) suffix
corresponds to the component presence in the resource and the \( m \) suffix corresponds to the component quality evaluation.

### 2.1 Verification of the measures

To verify the thesis that the structure of an e-learning resource conformant to the model of effective learning has a positive influence on the resource’s quality, we employed the above 36 measures to construct a questionnaire for evaluating e-learning resources (Stasiecka et al., 2005a, b). We used the questionnaire to analyze a population of 56 e-learning resources; they were given identification numbers from 1 to 56. The resources were of several kinds, for instance, virtual manuals, traditional books transferred to the virtual environment, courses without teachers, multimedia resources. Most of them concerned information technology, for example, programming languages and internet technologies. The respondents were teachers and students of technical universities.

We augmented the population of those resources with a pattern resource (against which we compare the above population’s elements); its identification number is 0. This pattern resource possesses all the components (level I) and sub-components (level II); they are of the best quality, which means that for each component, the presence measure has the value 1 and the quality measure has the value 5.

To evaluate the quality of a resource as a whole, we decided to use average marks calculated from the marks for the level II components, rather than the respondents’ subjective marks (where no identified measures are involved). The rationale is that the average marks seem to be much more credible because, as we observed, some marks for the level II components are missing in the respondents’ answers.

The data collected throughout the questionnaire was input into the GradeStat statistics program. GradeStat, which has been developed at the Institute of Computer Science, the Polish Academy of Sciences, performs a statistical analysis of multi-dimensional data; the results are presented in the form of various charts and maps. We decided to use the program due to its graphical illustration methods, especially the chart of ARs and overrepresentation map (Pleszczynska and Niewiadomska-Bugaj, 1999; Matyja, 2003; Kowalczyk et al., 2004; GradeStat, 2006).

In the first step, we used the GradeStat chart of ARs. AR is the name given (Kowalczyk et al., 2004) to the concentration index; it has a representation as an area contained in the unit square. AR’s value for a resource determines the extent to which the resource is dissimilar to the pattern resource for a given set of measures. The greater the [AR] the greater the dissimilarity between those two resources. For simplicity, in the paper we write AR instead of [AR].

Figure 1 shows the ARs for the considered population; the points on the OX axis correspond to the resources denoted by their identification numbers; the resources are grouped and ordered according to their average marks. The OY axis is for the values of the ARs.

In the chart, we can see a clear descending trend: the smaller the average AR in a group, the better the marks of a given resource. This observation leads to the conclusions that:

- the correct didactic structure of a resource has a positive influence on its quality; and
- the identified set of measures can be taken into further consideration to evaluate an e-learning resource quality.
2.2 Sufficient subset of measures
To determine which of the identified measures have the biggest influence on the dissimilarity between a resource and the pattern resource, we used GradeStat overrepresentation maps.

2.2.1 Overrepresentation map. An overrepresentation map presents dependencies between the elements of a given population (the map’s rows) and the measures describing those elements (the map’s columns) (Matyja, 2003; Kowalczyk et al., 2004; GradeStat, 2006). An example of an overrepresentation map is shown in Figure 2.
The height of a map’s row depends on the evaluation of the corresponding element against the entire population – elements of higher evaluation are illustrated with higher rows; similarly for the widths of the map’s columns. If the evaluation of a given measure is high compared to the evaluation of all the measures, then the corresponding column is wider.

Each field of the map represents a given measure for a given element of the population. Fields have various shades of gray depending on the comparison of the real value of the measure for this element to the measure’s expected value. The real value of the measure means the value given to the measure by the respondent. If the evaluation of an element is high compared to the evaluation of all the population elements, and the evaluation of a measure is high compared to the evaluation of all the measures, then the expected value of the given measure for this element is high. If the real value is equal to the expected value, we say that the measure for this element is neutral (the corresponding field is gray). If the real value is greater than the expected one, we say that the measure for this element is overrepresented (the field is black or dark gray). Finally, if the real value is less than the expected one, we say that the measure for this element is underrepresented (the field is light gray or white).

An important characteristic of the map is that the order of its rows and columns is determined by GradeStat (not by the user). Upper rows of the map represent those elements of the population, for which the measures corresponding to the left-most columns received high grades, and the measures corresponding to the right-most ones received low grades. Similarly, lower rows represent those elements of the population for which the measures corresponding to the left-most columns received low grades, and the measures corresponding to the right-most ones received high grades. If the set of measures describing a given population has been selected correctly (in the aspect of differentiating its elements), then the map is very similar to the ideal one; see Figure 2, where the darkest fields are put in the upper-left and the lower-right corners of the map. The other fields of the map have the following property: the farther from the diagonal towards the other corners of the map (the lower-left and the upper-right ones), the lighter the fields.

2.2.2 Finding a sufficient subset of measures with the help of the overrepresentation map. In the next step we decided to reduce the size of the identified set of the measures, that is, we decided to find the sufficient subset of those measures, which potentially can influence the dissimilarity between a resource and the pattern resource. To this end, we constructed an overrepresentation map for the population of e-learning resources presented in Section 2.1; see Figure 3:

- The rows of the map are labeled with pairs (resource identification number, resource average mark).
- The columns are labeled with pairs (measure identification number, suffix), where like previously, the \( p \) suffix corresponds to a component’s presence in the resource and the \( m \) suffix corresponds to the component quality evaluation by the respondent.

We can see on the map an interesting order of the resources and the measures. The upper rows represent rather good resources, while the lower rows represent rather bad resources. Analyzing the measures based on which dissimilarity was made (the left-most and the right-most columns), it is possible to indicate a subset of measures that can be used to differentiate good resources from bad ones:
Therefore, according to the correspondence between measures and components’ features, it is possible to indicate a subset of components that have the largest influence on the quality of a resource as a whole.
2.3 Weights

In the next step of our research, we focused on the weights corresponding to the importance of the level I and II components from the users’ point of view. In order to determine the weights, we constructed another questionnaire – the respondents were to assess the importance of each component by assigning them values (weights) from the set \{0, 1, 2, 3\}. The values had the following meanings:

- 0 – unimportant component; it has little influence on the resource’s quality.
- 1 – neutral component; it has some influence on the resource’s quality.
- 2 – important component; it should be included in the resource.
- 3 – very important component; it has to be included in the resource.

Similarly, in the case of our previous questionnaire, the respondents were teachers and students. This time there were 47 respondents: 21 teachers and 26 students.

In the beginning, we analyzed the population of all the respondents. The average weight of one component calculated for all the respondents’ answers was greater than 2 (precisely: 2.19), which means that according to the respondents all the components are important and should be included in a good quality resource.

Next, we subdivided the population into two subsets: teachers and students. The results of the analysis are shown in Figure 4. As we can observe, the importance of the presence and the quality of the components was higher for the teachers than for the students: the average weight for one component was 2.24 according to the teachers and 2.14 according to the students. For the teachers, the most important components were:

- 2. Main content.
- 2.2. New knowledge.
- 2.3. Examples of applying new knowledge in practice.
- 4.1. Self-evaluation.

For the students, the most important components were the following:

- 2. Main content.
- 2.2. New knowledge.
- 2.3. Examples of applying new knowledge in practice.
- 3.1. Recapitulation.
- 3.3. Dictionary of key concepts.
- 4.3. Feedback.

Furthermore, we can observe that the biggest differences for both subsets of the population were in the assessment of the importance for the following components:

- 3.1. Recapitulation.
- 3.3. Dictionary of key concepts.
- 3.4. Literature.
Figure 4.
The average values of the weights for the components of E-learning objects.
The outcome achieved for this simple students-teachers distinction proves that more useful results could be obtained if the population were subdivided according to other aspects, for example, base knowledge, age, motivation. We plan to do such research in the future – such aspects can influence the content and number of elements for a sufficient subset of measures used in the quality assessment activity.

3. A proposal of new metadata specifications

While analyzing the data collected by means of the first questionnaire, we concentrated on comparing the quality evaluation of e-learning resources based on the measures identified by us to the respondents' subjective quality evaluation. An interesting finding was that sometimes when a given resource was subjectively evaluated by the respondents as very good, it was not evaluated as very good when our measures were used.

Another interesting conclusion was that although resources with a correct didactic structure received high marks from the respondents, there were some resources with an incorrect didactic structure that also received high marks. This observation leads to the following question: what other elements (i.e. other than the didactic structure) did the respondents take into consideration when evaluating resource quality? The question is essential because, as we can observe, those elements made the incorrect structure of the resources less important than the other features (from the viewpoint of the respondents). In order to answer the question, we have developed an extended set of measures and a new version of our first questionnaire. The new set of measures was subdivided into the following groups:

- **Didactics.** It deals with estimating the level of conformance of the didactic structure of a resource to the model of effective learning.

- **Evaluation.** It concerns evaluating the results of the learning process. This group deals not only with testing and assessing the student, but also with assessing the usefulness and the quality of the resource itself. The following questions should help in assessing the quality of a resource in this aspect: what kinds of test questions does the resource involve? Can the questions be chosen by random? Can the history of previous sequences of questions and answers be taken into account when choosing new questions? Are there any open questions? Is there any feedback helping the student to see the right answer or to reach the right answer through some additional questions? Are there any reports available? Can the student evaluate the resource with regard to its quality, usefulness, etc?

- **Functionality.** It deals with determining whether the resource includes multimedia elements (that is, audio, video, graphics) supporting the student. We put forward some auxiliary questions: how was the quality of multimedia elements evaluated? Are there too many or too few such elements? When studying the resource, can the student use external databases or web services? Are there any interactive simulation elements, for instance, for studying programming languages? Are there any auxiliary elements/tools such as tables, dictionaries, calculators, etc? Are there any educational games?

- **Usability.** It concerns determining the ease of using a resource. The auxiliary questions are: can parts of the resource be imported from/exported to other environments? How is the presentation of the content evaluated? What kinds of navigation can be used? How are those kinds of navigation evaluated?
• **Reusability.** It deals with the reuse potential of a resource. The auxiliary questions are: does the resource include contact information (http, e-mail, etc.) to its authors? Is there any information about institutions/persons that recommend that resource? Is there any information about the number of its users (at present and in the past)? Is there any information about how the users evaluate the resource? Are there any keywords and brief information about the subject(s) covered by the resource? Is there any information about references to other relevant resources? Is there any information about the platforms on which the resource has been tested? Is there any information about how the resource conforms to the existing standards (are there any relevant certificates)?

• **Quality.** It concerns determining the quality of a resource as a whole, based on the characteristics of the above elements (from 1 to 5). Some other aspects are also considered.

According to us, to facilitate the potential user to determine the quality of an e-learning resource, the measures from the above groups should be included in metadata describing e-learning resources. Therefore, in the next section, we propose to modify some of the existing SCORM categories (especially Educational) and to introduce two completely new ones: Quality and Reusability.

### 4. Our proposal to extend the SCORM standard

Owing to the level of abstraction of our proposal, we give only the names and brief textual descriptions of the new metadata elements that we propose to define (Stasiecka et al., 2006b).

#### 4.1 Modification to the SCORM (educational) element

We propose to extend the SCORM Educational category by introducing the following child elements:

- (didactics)
- (evaluation)
- (functionality)
- (usability)

#### 4.2 New (reusability) element

The (reusability) element is a parent element; it contains the following child elements:

- (useContext) – represents the context(s) in which the resource can be used.
- (platform) – represents the platforms on which the resource has been tested.
- (standard) – represents the conformance of the resource to the standards.
- (contactInfo) – represents the (author, technical support, etc.) contact information necessary to use/reuse the resource.
- (reference) – represents the persons/institutions that recommend the resource.
- (user) – represents the persons/institutions that have used or are using the resource.
- (opinion) – represents opinions about the resource by its users.
4.3 New \textit{quality} element

The \textit{quality} element is a parent element; it contains the following child elements:

- \textit{basicQuality} – represents the quality of the resource with respect to the following categories: didactics, evaluation, functionality, usability, reusability.

- \textit{searchSupport} – represents the support to find information about the resource with the help of the resource’s key words and classification.

- \textit{userSatisfaction} – represents the quality of the resource with respect to its user satisfaction aspect.

- \textit{expertAppraisal} – represents the expert appraisal of the resource by independent experts and/or authorizing bodies.

- \textit{patternConformance} – represents the conformance degree of the resource to the ideal resource, that is, to the resource whose quality with respect to didactics, evaluation, functionality, usability, reusability, user satisfaction and expert appraisal are considered to be optimal.

5. Conclusions and future work

In the paper we have discussed our research on the issue of e-learning resources quality. We have identified and verified some set of measures; the choice of the measures was based on the conformance of e-learning resources to the model of effective learning. For this purpose, we used two questionnaires and the GradeStat statistics program with some interesting functionality to present graphically the dependencies for multi-dimensional data.

During the analysis we observed that there are also some other elements that have influence on the resource quality, different than those that we identified earlier (other than the conformance to the model of effective learning). These new elements should also be taken under consideration (e.g. evaluation, functionality, and usability).

One of the conclusions that we have reached is that the SCORM standard should be modified and extended. According to us, some of the currently existing SCORM categories should be modified (e.g. educational) and also two completely new ones (reusability and quality) should be introduced.

In our opinion, such a modification of the metadata is necessary, because due to the large and increasing number of e-learning resources, the issue of those aspects is becoming more important.

The proposal is still under research and needs more efforts. In particular, our future work concerns:

- further developing our concepts in order to incorporate them into some of the existing standards (especially, SCORM);

- defining in a precise manner: multiplicity, weights, and data types for the new metadata elements;

- further determining the sufficient subset of obligatory metadata that are necessary and sufficient to precisely describe various aspects of e-learning resources. For example, analyzing methods of subdividing users of e-learning
resources into various target groups to determine sufficient subsets of measures according to the needs of the identified groups; and
• developing an e-learning ontology.

For more details see Stasiecka et al. (2005a, b).

References

Corresponding author
Jacek Plodzien can be contacted at: jpl@ipipan.waw.pl

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints