INFORMATION QUALITY – A CHALLENGE FOR E-LEARNING 3.0

Doina Banciu

Institutul Naţional de Cercetare-Dezvoltare în Informatică, ICI Bucureşti
Doina.Banciu@ici.ro

Monica Florea

SIVECO România, Victoria Park, Bucureşti
Monica.Florea@siveco.ro

Abstract: The transition from the "*read*" web to the "*read/write"* web (e-Learning 2.0), and also to the "*read/write/collaborate*" web, which is the *e-learning 3.0* (*Edutainment*), the features of user-centric web, put the emphasis on the importance of quality information in e-learning systems. Many conceptual models to assess the information quality have been proposed: *A Conceptual Framework for Data Quality - CFDQ (Wang & Strong), Klein's model, PSP/IQ model*, etc. Starting from the Wang& Strong's model, Alkhattabi, Neagu and Cullen proposed a Framework *based on a Web-mining approach*. Despite the importance of quality evaluation of the content in e-learning systems, there is still work to be done in order to reach a consensus related to the methods and metrics for evaluation of quality information (IQ) and the evaluation standards to be used in e-learning 3.0 generation.

Keywords: e-learning, web 2.0, information quality, conceptual models

Rezumat: Tranziția de la „*Citeşte pe Web*" la „*Citeşte/Scrie pe Web*" (e-Learning 2.0), precum şi tranziția la „*Citeşte/Scrie/Colaborează pe Web*" (*e-Learning 3.0* sau *Edutainment*), caracteristicile web-ului centrat pe utilizator, pun accentul pe importanța calității informației în sistemele de e-Learning. Au fost propuse numeroase modele conceptuale de evaluare a calității informației: *Model conceptual pentru Calitatea Date*lor - *CFDQ* (Wang & Strong), *Modelul lui Klein*, *Modelul PSP/IQ,* etc. Pornind de la modelul lui Wang & Strong, Alkhattabi, Neagu şi Cullen au propus un model bazat pe explorarea Web. În pofida importanței evaluării calității conținutului în sistemele de e-learning, nu s-a ajuns la un consens privitor la metodele şi metricile de evaluare a calității informației (IQ) şi a standardelor de evaluare utilizate în generația e-learning 3.0.

Cuvinte cheie: e-learning, web 2.0, calitatea informației, modele conceptuale

1. Introduction

The development of internet technologies starting with the '90s created the premises for the development of e-learning services. The evolution of these services encompasses the following stages: Web-based learning, computer-based learning, online learning, e-learning, distance learning. The definitions of e-learning varied from web-based self-study to real-time learning and collaboration. Mainly, eLearning refers to the use of internet or wireless technologies to deliver a broad array of training solutions. Marc Rosenberg (2001) [12] suggested the following definition of eLearning: "*the use of Internet technologies to deliver a broad array of solutions that enhance knowledge and performance*". According to The eLearning Action plan (2001): Designing tomorrow's education: "*e-Learning is the use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services as well as remote exchanges and collaboration*". Even though there are different meanings of e-learning to almost everyone who use it, almost all agree that eLearning is of strategic importance.

2. e-Learning Generations

The first model of e-learning was called also e-Learning 1.0. Starting with Web 2.0 technologies -*Web as a platform* was first called *web 2.0* by Tim O'Reilly in 2004 [11], the second generation of web development and design, aiming to support and improve the communication, information exchange, information security, interoperability and collaboration on World Wide Web, a new e-learning model arises: e-Learning 2.0.

Anderson [2] describes six concepts behind Web 2.0:

- Individual production and User Generated Content
- Harness the power of the crowd
- Data on an epic scale
- Architecture of Participation
- Network Effects
- Openness

As Downes mentioned [4]: WEB *2.0 is an attitude, not a technology*, meaning that we are not facing a technological revolution, but a social revolution.

e-Learning 2.0 uses *web 2.0 technologies* for educational purposes, the most common features are:

- *social networks*;
- *wikis:* a webpage or set of web pages that can be easily edited by anyone who is allowed access;
- chats;
- *blogs:* simple webpage consisting of brief paragraphs of opinion, information or links, called *posts*, arranged chronologically;
- *RSS reader pages*;
- *social bookmarking*: allow users to create lists of 'bookmarks' or 'favorites', to store these centrally on a remote service);
- *multimedia-sharing*: facilitate the storage and sharing of multimedia content;
- *folksonomy*: a collection of tags created by an individual for their own personal use;
- *collabulary* :collective vocabulary;
- *group work spaces* etc.

e-Learning 2.0 is based on the following concepts [16]:

- standards and technology;
- sharing local and external resources;
- process and governance;
- organization culture and education;
- abilities and competences.

A new generation is coming: *e-Learning 3.0*, being called also *Edutainment* or *Entertainment-Education*. E-Learning 3.0 [17] is based on:

- *cloud computing*: distributed computing, increased data storage and retrieval, easy access to tools and services that enable personalized learning, self learning;
- *collaborative Learning*: e-Learning 3.0 will facilitate collaborative learning through predictive intelligent filtering, intelligent agents, multi-user participative features;
- *3D visualization and interaction*: the development of 3D multi-touch interfaces and multi-gesture devices will facilitate exploration of virtual spaces, manipulation of virtual objects, fine motor skills interaction;
- *mobile intelligent technologies*: the extension of intelligent mobile technologies will play an important role in e-Learning 3.0. Using smart phones and better connected network services (wireless, satellite), education and learning will be accessible to

learners anytime and anywhere.

While e-learning 1.0 represents the "*read phase*" web, e-learning 2.0 represents the *"read/write" web*, the web being transformed into a user-centric web. E-learning 3.0 will make the shift to *"read/write/collaborate" web*.

Among the advantages of e-learning systems versus traditional learning, we can mention:

- accessibility and mobility;
- flexibility;
- adaptability;
- new dynamic technologies;
- reduced costs;
- the courses are organized on the subjects / topics, not on the age groups like traditional courses.

We have noticed some interrelated transformations in the e-learning field over the next few years [12]:

- e-learning becomes more than "e-training";
- e-learning moves to the workplace;
- blended learning is redefined (combining formal training with non-formal training);
- e-learning is less course-centric and more knowledge-centric;
- e-learning adapts differently to different levels of mastery;
- technology becomes a secondary issue.

3. Assessing Quality in e-Learning Platforms

"E-learning platforms" is a term covering a variety of different products, all of which support learning in some way by using electronic media [5].

An e-learning platform is a system for management of content and teaching, being made of several components:

- Learning Management System (LMS);
- Learning Content Management System (LCMS);
- Human and Knowledge Management System (HKMS),
- Virtual class (synchrone learning);
- ePortfolio.

E-learning systems' quality is made of the following components (Garvin, 1988) [6]:

- performance – the e-learning system should perform in a efficient manner, taking in consideration the user's requirements;
- functionalities according to the requirements;
- reliability;
- conformity with the standards (technological, industrial, and educational standards);
- durability: the e-learning system should be relevant also from pedagogical perspective, and easy to be updated;
- flexibility: the system should be easy to be repaired and adjusted according to the requirements;

- aesthetic;
- perceived quality.

Due to their multidisciplinary nature, the assessment of e-learning platforms requires the collaboration of experts from different domains like: computer science, information systems, psychology, education, educational technology, etc.

The evaluation of e-learning platforms requires a measure of self review of the institution on the following areas [18]:

- institutional cultural change towards the adoption of eLearning platform;
- content management: identification, storage and retrieval of digital content;
- communication and collaboration (inside and outside the institution);
- learner information: assure an effective management of learner data;
- administration: network access and users and groups management;
- ICT resources: security, access, administration rights, methods of data storage, filtering etc.

In the literature, there are defined a series of conceptual models (frameworks) to assess the quality of e-learning systems. These conceptual models can be grouped according with the e-learning systems' aspects that are envisaged:

- models to assess the quality of the content (information);
- models to assess the quality of the e-learning services;
- models to assess the quality of the e-learning programs;
- models to assess the quality of the e-learning institutions,
- models to evaluate the acceptance of e-learning technologies (extending the general conceptual models of technologies' acceptance to e-learning technologies);
- models for evaluation of e-learning systems student-centric;
- models for evaluation of e-learning systems using standards specifications (e.g. SCORM specifications).

In the next section we will make a review of the models to assess the quality of the information (IQ) in e-learning platforms.

4. Models to Assess the Information Quality in e-Learning Platforms

4.1 Information Life Cycle (POSMAD) is Made of the Following Steps [10]:

- **Plan**—Identify objectives, plan information architecture, and develop standards and definitions; many activities associated with modeling, designing, and developing applications, databases, processes, organizations, and the like;
- **Obtain** - Data or information is acquired in some way; for example, by creating records, purchasing data, or loading external files;
- **Store and Share** - Data are stored and made available for use;
- **Maintain** - Update, change, manipulate data; transform data, match and merge records etc.;
- **Apply** - Retrieve data; use information. Includes all information usage such as completing a transaction, writing a report, making a management decision, and completing automated processes;
- **Dispose** - Archive information or delete data or records.

Four key components affect information quality in organizations:

- **Data (What)** - Known facts or other items of interest to the business;
- **Processes (How)** - Functions, activities, actions, tasks, or procedures that manipulate the data or information (business processes, data management processes, processes external to the company, etc.);
- **People and Organizations (Who)** - Organizations, teams, roles, responsibilities;
- **Technology (How)** - Forms, applications, databases, files, programs, code, or media that store, share, or manipulate the data, are involved with the processes, or are used by the people and organizations.

4.2 A Conceptual Framework for Data Quality - CFDQ (Wang & Strong, 1996)

Wang and Strong [15] developed a conceptual framework for data quality, organizing the quality information dimensions (15 dimensions) in four categories: *Intrinsic quality, Accessibility, Contextual quality, Representational quality.*

Table 1. Wang and Strong Quality Categories and Dimensions

Quality categories	**Dimensions**
Intrinsic quality	Accuracy
	Objectivity
	Believability
	Reputation
Accessibility	Accessibility
	Security
Contextual quality	Relevancy
	Value-Added
	Timeliness
	Completeness
	Amount of Information
Representational quality	Interpretability
	Ease of Understanding
	Concise Representation
	Consistent Representation

4.3. In 2002, **Klein** [8] depicted 5 dimensions from Wang & Strong's 15 dimensions, and focused on the detection of the factors that are associated with incidents in which users detect information quality problems on the World Wide Web. Also he tried to figure out if users who have experienced information quality problems have different perceptions of the quality of information available on the

World Wide Web than users who have not experienced problems:

Table 2. Preliminary Factors Associated with Information Quality Problems on the World Wide Web

Quality dimension	**Preliminary factors**
Accuracy	Discrepancy Timeliness Source/Author Bias/Intentionally False Information
Completeness	Lack of Depth Technical Problems Missing Desired Information Incomplete When Compared with Other Sites Lack of Breadth
Relevance	Irrelevant Hits When Searching Bias Too Broad Purpose of Web Site
Timeliness	Information is Not Current Technical Problems Publication Date is Unknown
Amount of Data	Too Much Information Too Little Information Information Unavailable

4.4. Based on CFDQ model of Wang & Strong, Kahn et al. [7] defined the **PSP/IQ model** (Product and Service Performance / Information Quality) for product/service quality (2 quality types) dividing the 16 dimensions of quality information in 4 categories (*Sound information, Useful information, Dependable information, Useable information*):

Table 3. PSP/IQ Model

Quality type	**Classification**	**Dimension**
Product quality	**Sound information** The characteristics of the information supplied meets IQ standards	Free-of-Error Concise Representation Completeness Consistent Representation
	Useful information The information supplied meets information consumer's needs.	Appropriate Amount Relevancy Understandability Interpretability Objectivity
Service quality	**Dependable Information**	Timeliness Security

	The process of transforming data into information meets standards	
	Useable Information The process of transforming data into information exceeds information consumer needs	Believability Accessibility Ease of Manipulation Reputation Value-Added

4.5 Framework Based on a Web-mining Approach

Based on the Wang & Strong framework, Alkhattabi, et al.[1] proposed a framework for measuring the information quality in the e-learning systems, consisting of 14 quality dimensions (attributes) grouped in three quality factors: *intrinsic, contextual representation, and accesibility*.

Intrinsic Quality

Quality attributes:

- **Objectivity:** Published materials should provide impartial information

Questions	**Metrics**
There is information on the publisher organization?	X = 1 (if there is information), X = 0 (if there is not information)
The domain extension refers to unbiased bodies and is appropriate for the content?	X = 1 (if it is relevant for the content), X = 0 (if it is not relevant)
What is the average of links in the system that refer to unbiased bodies?	X= UL/TL, $0 \leq X \leq 1$ UL = number of unbiased links TL = total number of links

- **Accuracy:** Published materials should provide accurate information:

Questions	**Metrics**
There are references provided in order to check how accurate is the published material	X = 1 (if there are references), X = 0 (if there are not references)
Is the page that contains the information dated?	Last update in the system is mentioned? X = 1 (if yes), X = 0 (if no)
Are there additional links for further reading and suggested resources?	Information contains any additional resources? X = 1 (if yes), X = 0 (if not)

- **Believability:** Published materials should provide believable information

Questions	**Metrics**
Wo wrote the information?	Is there any information on the author? X = 1 (if yes), X = 0 (if no)
Wo publish the information?	Is there any information about the publisher body? X = 1 (if yes), X = 0 (if no)
Can we contact the author or the organization?	Is there any contact information X = 1 (if yes), X = 0 (if no)

Accessibility – quality attributes

- **Availability:** The availability of information resources in the e-Learning when needed

Question	**Metrics**
What is the current availibility rate for the system?	$X= (P_h –O_h)/ P_h$, $0\leq X \leq 1$ P_h = prime hours O_h = outage hours

- **Relevency**: The provided information should be relevant to the topic under study

Question	**Metrics**
When querying using a reputable search engine, the systems appears in the first ten results when searching for sites related to the subject under study?	$X = (\sum SP_i)/k$, $1\leq i\leq k$, $0\leq X \leq 1$ SP_i= the probability to have the system within the first 10 results when querying the search engine (Google) for the keyword i K = total number of needed keywords

- **Accessibility:** All the elements of provided information should be accessible for all users

Questions	**Metrics**
What proportion of components (links, files, media etc) can be accessed by users?	X = 1 – NC/Com, $0\leq X \leq 1$ NC = number of inaccesible components (broken links etc) Com = total number of components

- **Response time:** The waiting time for the system to response to a specific task

Questions	Metrics
How long does it take before the system response to a specific request?	X = TT – ST, X>0 TT = total time to complete the request ST = time of submitting the request

Contextual representation - quality attributes

- **Conciseness:** Provided information is concise

Questions	Metrics
What is the extend of using hierarchical branching for information about provided keywords in the system?	X= LK/PK, $0 \leq X \leq 1$ LK= number of keywords with links PK = number of provided keywords

- **Verifiability**: Provided information can be checked for correctness

Questions	Metrics
Is there enough references for each key concept in the system?	X= RK/PK, $0 \leq X \leq 1$ RK= number of keywords with at least one reference PK = number of provided keywords

- **Representational consistency**: Provided information is represented in a consistent manner

Questions	Metrics
How consistent is the representation of provided information in the system?	X = 1- D/P, $0 \leq X \leq 1$ D = pages in the system with different style sheets P = total number of pages. Home page will be used as a reference

- **Understandability**: Provided information can be easily understood

Questions	Metrics
What proportion of explanation components (examples, figures, etc) can users access?	$X = PE_x/P$, $0 \leq X \leq 1$ PE_x = pages in the system with explanation components P = total number of pages

- **Amount of information**: The amount of provided information is appropriate for the task

Questions	**Metrics**
Is the amount of provided information too much or too little?	$X = \|Z\|$, $X \geq 0$ $Z = 1 - PK/NK$ PK = number of provided keywords NK = number of needed keywords If $Z>0$, too little information If $Z<0$, too much information

- **Reputation:** The web impact factor for the choosen system: how the system has been judged in general as information source

Questions	**Metrics**
What is web impact factor (WIF) for the system?	$X = LP/I_nP$, $X \geq 0$ LP = total link pages I_nP = number of web pages published in the website which are indexed by the search engine (Google)

- **Completeness**: The available resources have all the needed information

Questions	**Metrics**
How complete is the provided information?	$X = 1 - MK/K$, $0 \leq X \leq 1$ MK = number of missing keywords K = total number of needed keywords

For each quality attribute, each defined metric will be assigned a direct score between 0 and 1. For each metric, a higher value reflects a better quality score in terms of the corresponding attribute. The quality score for each attribute is calculated as the average of the values of the related metrics.

To compute the overall quality score, this study will use the assigned relative importance weight for each attribute within the main quality factors, and a relative importance weight for each factor in the overall quality within the proposed framework. The following equations will be used:

- To calculate the quality score for each quality attribute: $X_{j,i}= AVG(M_{j,i,k})$, $k \geq 1$, where $M_{j,i,k}$ represents the quality metrics corresponding to the quality attribute i inside the main quality factor j.

- To calculate the quality score for each factor: factor Quality Score $(Y_j) = \sum \alpha_{j,i} X_{j,i}$, where $\alpha_{j,i}$ represents the relative importance of the quality attributei inside the main quality factor j, and $X_{j,i}$ represents the quality score given to the same attribute.

- To calculate the overall quality score: overall Quality Score $\sum \beta_j Y_j$, $j=1,3$, where β_j represents the relative importance of the factor j in the overall quality, and Y_j represents the quality score given to the same factor.

The proposed framework can be used to evaluate, compare, and rank information quality in e-learning systems.

4.6 Adapting Quality Frameworks

The evaluation of the information in e-learning systems becomes more complex due to the development of Web 2.0 technologies that are imbedded in new generation of e-learning. In the context of user-centric web, e-learning 3.0 faces the great challenge of adapting the evaluation standards, methods of evaluation and metrics to the features of new education era.

The proposed quality frameworks cannot be applied directly in new education environement, they need to be mixed and adapted to the context.

We have to take into consideration the evaluation of the features of e-learning 3.0:

- content relevance according with the user type and skills (e.g. the level of IT skills: regular user or specialist in birotics);
- personalization of the content;
- interactivity;
- novel, humorous and relevant content: serious games, role playing, simulations, augmented reality;
- sensory appeal: the content appeals to multiple senses: audio modules, graphics, animation, video modules;
- adaptability of the content (e.g. for the users with special needs, to show a text instead of a image);
- Intellectual Property Rights of the digital content.

We propose to further expand the Wang & Strong framework by adding more quality dimensions, using Alkhattabi approach of giving scores to quality dimensions and also prioritize the dimensions (associating weights) according with their importance in a specific context. Also adequate metrics are required in order to assess all the features of e-learning 3.0.

5. Conclusions

The user –centric web, the shift from a "read" web to "read/write" web (e-learning 2.0), and even more to a "read/write/collaborate" web, raised an important question for the evaluation of information quality in e-learning systems. Although e-learning services and the quality of content in e-learning systems becomes of high importance nowadays, there is still work to be done in order to reach a consensus related to evaluation standards and methods of evaluation of information quality in such systems. One of the challenges of e-learning 3.0 is the assessment and assuring the quality in e-learning systems.

REFERENCES

1. **ALKHATTABI, M.; NEAGU, D.; CULLEN, A.:** Assessing information quality of e-learning systems : a web mining approach. Computers in Human Behavior 27, 2011, pp. 862-873.
2. **ANDERSON, P.:** What is Web 2.0? Ideas, technologies and implications for education. JISC reports. http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf, 2007.
3. **BARTOLOMÉ, A.:** Web 2.0 and new Learning paradigms, eLearning Papers, No. 8, April 2008, ISSN 1887-1542, 2008.
4. **DOWNES, S.:** e-learning 2.0, 2006.
5. **FERL:** Evaluating Learning Platforms, 2005.
6. **GARVIN, D.:** Managing quality. New York: Macmillan, 1988.

7. **KAHN, B. K.; STRONG, D. M.; WANG, R. Y.: I**nformation quality benchmarks: product and service performance. Communications of the ACM, vol. 45, no. 4, pp. 184-192, 2002.
8. **KLEIN, B. D.:** When do users detect information quality problems on the World Wide Web? American Conference in Information Systems, 2002.
9. **KNIGHT, S. A.; BURN, J. M.:** Developing a Framework for Assessing Information Quality on the World Wide Web. Informing Science vol.8, pp. 159-172, 2005.
10. **McGILVRAY, D.:** Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information, Morgan Kaufmann Publishers, 2008.
11. **O'REILLY, T.:** What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, Available: http://mpra.ub.uni-muenchen.de/4580/, 2007.
12. **ROSENBERG, M. J.:** Beyond E-Learning: Approaches and Technologies to Enhance Organizational Knowledge, Learning and Performance, 2006.
13. http://www.marcrosenberg.com/images/What_Lies_Beyond_E-Learning_ASTD.pdf
14. **SOUMPLIS, A.; KOULOCHERI, E.; XENOS, M.:** The Twobility Factor, The 7th International Conference eLearning and Software for Education, Bucharest, 2011.
15. **WANG, R. Y.; STRONG, D. M.:** Beyond accuracy: what data quality means to data Consumers. Journal of Management Information Systems vol. 12, no. 4, 1996, pp. 5-34.
16. http://steve-wheeler.blogspot.com/2009/04/learning-30.html
17. http://www.debaird.net/blendededunet/2007/02/learning_30_mob.html
18. http://www.excellencegateway.org.uk/page.aspx?o=ferl.aclearn.page.id461