Is e-learning ready for big data? And how big data would be useful to e-learning ?

Paolo Maresca[1], Andrea Molinari[2,3]

[1]Department of Ingegneria Elettrica e Tecnologie delle Informazioni (DIETI) – Università Federico II di Napoli (ITALY)
[2]School of Industrial Engineering and Management, Lappeenranta University of Technology (FINLAND)
[3] Department of Industrial Engineering - University of Trento (ITALY)

Abstract **The paper presents an overview of possible application fields of big data to the Technology-Enhanced Learning (TEL), with the many different facets this could imply. Many are the benefits for e-learning when approaching the collection of data, especially when e-learning is delivered in a life-long learning perspective. All these benefits could impact the future of eLearning, by revolutionizing the way we analyze and assess the eLearning experience. On one side, we present our experience in enriching the persistence layer of an LMS with a deeper log system on users' actions, in the perspective of collecting volumes of data compatible with big data tools and technologies, while highlighting some related issues. On the other hand we will deal with the first applications of cognitive systems that are responsible for catalysing the big data in analytics aimed at e-learning activities.**

Keywords: e-learning, big data, LMS architecture, cognitive systems

I. INTRODUCTION

Today, big data is one of the buzzwords that IT researchers and specialists use everywhere. So we have seen various "waves" of attention and "hype" on technologies and solutions related to Technology Enhanced Learning, but many of these hypes have revealed to be simply another buzzword and have not been consolidated. Today's buzzword that has been sweeping the world for a few years but has only just started appearing more commonly in eLearning is Big Data [10].

The term "Big data" refers to the huge amount of data coming from many different data sources, that become too large, complex and dynamic for any conventional data tools to capture, store, manage and analyse. Big data approaches and technologies interest many different application fields, and we will find "big data …." in front of any data analytics tool. Most of the problems were related with data warehouses, technologies aimed at supporting decisions based on reconciliation of the different databases of the organization.

The research activities in big data are aimed to find faster and more scalable solutions to store and process all data collected, instead of using traditional data warehousing approaches that are expensive, hard to design and to implement. Big Data therefore introduced two issues: how to address the problem of storing such a large amount of data, and how analytics tools could be created for the problem of analysing these huge datasets.

The paper will present the experience of the authors in designing and implementing a mechanism that will generate flushes of data from the persistence of the virtual learning platform created for managing tens of thousands users in our region, both from Academy and Industry. After the design of the solution, we have implemented a traditional internal method for managing the first data source for big data, i.e., dedicated logs of the platform. With a potential of more than 70.000 users, the volume at the moment are still in the range of the high level of database application, with approximately ten of thousand users using every day the platform for their e-learning tasks.

The tasks are recorded both for educational and security reasons, being the logging tasks involved in the recording of mouse clicking and in SCORM(Shareable Content Object Reference Model)-based material logging. Being SCORM logging not enough for some educational paths, according to the known limitations of SCORM standard, we decided to implement a meta-SCORM service, where more SCORM packages could be used in an educational path. In this scenario, we have logs of the platform for clicks and users' actions, logs from the SCORM player, and logs from the educational path services.

In general, there are several elements of data gathering and manipulation inside e-learning platforms [11] that could push this application field towards tools and techniques typical of Big Data:

- Traditional Weblogs, being the application a web-based software;

- internal logs of usage of the platform, the so-called "digital breadcrumbs", that track the learner's journey throughout the entire learning experience;

- Mobile logs, where data about mobile learning actions are collected;

- Service logs, users' actions on the different elements of the platform like documents, forums, blogs, FAQ etc.;

- logs from the SCORM player, normally an external entity respect to the core services of

the platform, with the records of the SCORM objects' execution;

- Tin-CAN API calls, in case the platform is connected or acting as a Learning Record Store (LRS).

Recently, a further set of new ideas in e-learning could increase the need of a structural change of Learning Management System (LMS) architecture towards approaches and technologies connected with big data:

- Massive Open Online courses (MOOCs), by definition a generator of high volumes of data

- Life-long learning, an old buzzword of e-learning that is still valid and interesting and, most of all, is another generator of big data, specifically along time;

- Serious games that will use materials inside the platform, thus generating a relevant dataset related with users' performances;

- Big Data will change the way we approach e-Learning design by enabling developers to personalize courses to fit their learners' individual needs [14].

The paper will present an overview of problems and opportunities related with the introduction of a big data approach to e-learning, both in the software architecture of the platforms and in the approach of e-learning stakeholders to this discipline. The paper is divided as follows: section 2 will present the potential application fields of big data technologies and approaches in e-learning. Section 3 will illustrate one specific example of big data in e-learning, i.e., the change in the persistence layer of a software platform for e-learning, while section 4 will present the use of big data technologies, specifically IBM "ask Watson" solution to e-learning settings.

II. BIG DATA AND E-LEARNING: THE POTENTIAL APPLICATION FIELDS

There are a number of reasons why big data may, very well, revolutionize the eLearning industry. First and foremost, it will allow eLearning Professionals to customize the eLearning experience to provide learners with more effective, engaging, and informative eLearning courses and modules. Big data also has the potential to impact the future of eLearning by:

a) Offering invaluable feedback.

While online surveys and discussions may offer feedback regarding the effectiveness of eLearning courses and modules, big data gives to eLearning professionals the chance to receive invaluable feedback that can be used to pinpoint where the learner, and the eLearning course itself, may need to be improved. For instance, if a learner is able to look at an analysis of where he/she fell short while taking the eLearning course, he/she can then figure out how to correct the issue moving forward. At the same time, if the online facilitator observes that the majority of the learners struggle with a particular module or assignment, he/she can make proper adjustments to improve learners' performance.

b) Allowing eLearning professionals to design more personalized eLearning courses.

If eLearning professionals are given the opportunity to know how their learners are acquiring information and what works best for them, in terms of content and delivery, then this will result in more personalized and engaging eLearning courses. As such, modules can be custom tailored to meet the individual needs of the learner, which will offer a high quality and meaningful learning experience.

c) Targeting effective eLearning strategies and eLearning goals.

Big data in eLearning gives us an inside look at which eLearning strategies are working and which ones aren't necessarily helpful in terms of eLearning goal achievement. For example, you can determine which eLearning courses are contributing to skill development and which eLearning modules or elements may be irrelevant. As such, you can then devote resources to the aspects that are working, so that the learners can receive the preparation they need to fulfill their career goals.

d) Tracking learner patterns.

With big data, eLearning professionals gain the rare ability to track a learner throughout the entire process, from start to finish. In other words, you can see how well they performed on a test, or how quickly they finished a challenging eLearning module. This helps you to pinpoint patterns that will not only enable you to learn more about the learning behaviors of the individual learner, but your learners' group as a whole.

e) Expanding our understanding of the eLearning process.

As eLearning professionals, it's essential that we learn as much as possible about how learners acquire and digest knowledge. Big data gives us the chance to gain an in depth understanding of the eLearning process and how the learners are responding to the eLearning courses we are delivering to them. We can even figure out which time of day they learn most effectively or which delivery methods allow them to retain information most efficiently.

This information can then be used to take our eLearning strategies to the next level. To make a long story short, embracing Big Data reshapes eLearning and it's a win-win situation for both the eLearning professionals and the learners.

III. One Testbed For Big Data Generation: The "Online Communities" Platform

The source of inspiration for big data analysis in e-learning is a virtual communities management platform entirely and autonomously created by our research team, starting since 1998. The approach used in most of the communities managed by the platform regards what is called "blended" approach, i.e. an e-learning mixed between frontal and online education training, asynchronous and synchronous, with online tutoring and frontal work sessions, all supported by our "Online Communities" system, of which, around the end of the 90s, the Faculty of Economics of the University of Trento has decided to adopt, followed by other public and private institutions.

Currently, "Online Communities" (OC from now on) is mainly used outside the university campus, serving approximately 50.000 users from different public and private customers against approximately 15.000 students in our University [16].

OC is a dynamic web application, based on the metaphor of the virtual learning community, which ensures the cooperative organization of work in groups of users called "Community". A virtual community is defined as a space of communication shared by a group of people, not only related to educational aspects. Every community has at least one coordinator and participants that are not anonymous. It is natural to imagine a virtual community as an aggregation of individuals made possible thanks to computers; an extension in the virtual learning environments is the class in which the courses take place.

The system is designed from the ground up within the Laboratory of Maieutic working group - Department of Industrial Engineering of the University of Trento, and is able to support the needs of a broad group of users (teacher, student, tutor, lecturer, external consultants, supervisor, dean, counsellor, secretary, board member etc.), customizable within the context in which the system is used (for example in a business organization we will have different roles respect to the university, as president, secretary, director, administrative, board of director etc.).

The participants in the system are not anonymous, and have a number of roles; each role brings with it specific rights and duties. Therefore, the enrolled users participate in a series of communities, fulfilling different roles in each of them. The communities are also characterized by a series of events that correspond to the active involvement of members in different moments. OC was released outside of test environments in 1998 as a working prototype, then reached through different evolution steps its maturity in early 2005, but counting on a long experimentation on a limited number of courses started at the end of 1998.

Since September 2005, the system is in operation at the University of Trento, involving tens of thousands of users, and has become an everyday tool for the teaching of many teachers. It represents the technology infrastructure used officially by some faculties of the university of Trento for testing and supporting new forms of teaching based on the techniques and the methods of e-learning. In 2007, the Autonomous Province of Trento decided to adopt it as its platform to deliver e-learning for its approx. 20.000 users. Then the Chamber of Commerce of Trento and the Chamber of Commerce of Bolzano adopted the platform and extended it to the affiliated enterprises, where we have just for the Province of Trento approx. 55.000 enterprises with the respective users.

The platform enables the members of any community to collaborate using multiple computer services (blog, wiki, chat, video conferencing, forums, file sharing, diaries , diaries, etc.). These services are used in areas collaboratively, both in the training, given the apparent closeness between the two environments. It 'also clear that, talking about continuing education and lifelong learning, the boundary between training and collaboration is increasingly being blurred when dealing with adults involved in the workplace. The logic of a virtual community platform considers any combination of people, brought together in a virtual place (hence "virtual community") for various purposes.

From our analysis, at the moment the interaction of users generate 2GB of data per month only for the users' actions log, just for 2.000 users/day. It's clear that the overall picture could become much more compelling for any software platform where the collection of data could stimulate and support DSS from the top management of the institution, or where hundreds of thousands participants follow intensive MOOCs.

Most of this data are coming from the first data source we used in our experimentation due to its affinity with big data data sources, the aforementioned "Actions". It collects all data coming from users' interactions with any OC object or service. It is a sort of "sensor" introduced inside the source code of the platform in any place is needed the software to capture an "action" from the user interface. This is of course a relevant enrichment of the logs recorded by the web application server, and has been used for many different purposes.

Due to volume issues, the system at the moment is blocked on collecting only some types of events, to a certain granularity defined by the system administrator. This choice has not been a design choice, but a performance-related one. From the early experimentations, it was clear that the amount of data could have compromised the capacity of the DBMS to stand data acquisition pace and volumes: a typical "Velocity" and "Volume" big data problem.

Tasks are recorded both for educational and security reasons, being the logging tasks involved in the recording of mouse clicks and SCORM-based material actions. Being SCORM logging not enough for some educational paths, according to the known limitations of SCORM standard, we decided to implement a meta-SCORM service, where more SCORM packages could be used in an educational path, mixing them with other services provided by the platform (like Wikis, FAQs, forums etc.) in a unique view called "educational path". In this scenario, we have logs of the platform for clicks and users' actions, logs from the SCORM player, and logs from the educational path service.

From the potential amount of data generated by these actions, we started to imagine our big data approach to the platform.
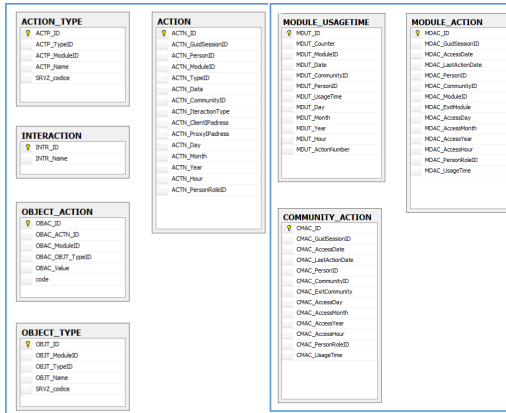

Figure 1. data structures for big data analysis in "Online Communities"

IV. BIG DATA AS SUPPORT TO LEARNING MANAGEMEMENT AND LEARNING PROCESS

One of the domains in which you can take on the challenge of big data and the cloud is the *management of learning processe*s as well as the *customization of the learning managemen*t. The *customization of learning processe*s can be accomplished employing the cognitive system as the IBM Watson [2] according to a paradigm known as to **ask Watson,** with which you establish a triage *Watson-student-teacher* that will be explained in detail in the following paragraphs.

Thanks to the use of cognitive computing, classes become more intelligent, not only for the content they make available to the students but for the possibility of using student's data to figure out how to grow the class. The observation period may be as long as much as they are in the school and beyond[3].

These systems can also help teachers deal with the problems of the students such as school dropout, credits recovery etc., thus allowing a process that aim at learning *customization*. The paradigm is innovative and is based on the fact that the teacher asks the cognitive system to "talk" with the student in order to understand which are the strengths and weaknesses to be reported to the teacher who then will decide which course materials should be delivered to overcome the identified learning problem. This feature is known as **Ask to Watson** and represents the aforementioned **triage** *student-Watson-teacher* which represents the true individual learning customization.

As an example, you can **ask Watson** to discover the learning mode of the student (kinaesthetic, visual, auditory) in order to deliver the appropriate training material. Or you can **ask Watson** "to prepare a small subclass for the Olympics of mathematics or computer science" and then provide the material and proper training to achieve particular objectives challengers. But

the most important thing is the interaction with Watson that occurs in spoken language.

To date, Watson interacts in English, Japanese, Spanish and Portuguese, it learns from the students using the features outlined in the previous paragraph. In addition, the teaching material is densely developed with HTML tags and hypertext, so that you do not need to do complex searches and the concepts are often proposed using games and flowing into the domain appropriate to young adults.

There are many existing projects in this area around the world. Georgia's Gwinnet County public schools, is the 14th most populous district of the USA and is one of the first college that has pioneered the use of Watson. Watson has the task of "*identifying similarities in how students learn and predicting performance and learning needs",* while "*specific content and teaching techniques can be aligned to each of the district's 170,000 students to Ensure the best learning experience"*. In particular, the paradigm is quite complex and is constituted by recursive cycles that consist of the following 6 steps:

i) Watson identifies weaknesses and strengths of every students;
ii) Watson recommends behaviour and contents for students aligned to their skills and learning styles;
iii) Teacher selects appropriate learning path and Creates a progression plan for student;
iv) Students consume the recommended content from the material plan;
v) Teachers monitor the progress of the students and adjust the plan;
vi) Teachers use Watson to identify student skills attainment aligned to the standards defined.

More specifically, it is interesting to find out how they interact between Watson, the student and the teacher through the services it offers. The interaction occurs in spoken language and consists of three cognitive services: I) **Ask**, II) **discover**, III) **decide** delivered in sequence. In the first the student asks Watson to guide him, in the second Watson teaches the student how to derive the answers from the questions asked, in the third the teacher sees the results and delegates Watson for the adaptation of the training course. This mode of *assistance* is very promising and is bearing much fruit.

Other challenges have been launched in this field especially in the analysis of data and analytics [1][6][7]. For example, Wichita State University uses advanced analytics to predict potential students chances of success: 15% boost in registration. Hamilton County, Tennessee's Department of Education uses predictive analytics to improve student achievement causing the Following results: 8% increase in the graduation rate to 80% and 25% reduction in the annual drop rate. Seton Hall has used integrated marketing optimization solutions analysing social media to understand how students move through the recruitment process: 13% enrolled increase.

The novelty of this approach lies in the use of natural language (English, Spanish, Portuguese and Japanese) that allows Watson to discover the student's difficulties

and to report them to the teacher who chooses which educational materials to provide, and so on until you reach the training course objectives that have been established for that student. Training materials can be chosen on the basis of intellectual channels of young people and can be provided on the basis of the receptive attitudes of the students (auditory, visual or kinaesthetic).

At this stage, the interaction through natural language helps a lot especially if you are using two essential functions of cognitive system that are called speech to text and text to speech (see fig.2). The first transforms written text into speech, the second performs the inverse function. Based on these features, for example Sundararajan and Nitta [12], designed and realized a tutoring system for K12 students, intensively using interactivity, automatic generation of questions, and learning analytics. It is worth noting that these features, as all the others, are available through Platform-as-a-Service (PaaS) bluemix IBM and the access for Universities is a free of use under the IBM academic initiative agreement for six months.

Watson Ecosystem is one of the largest research organizations in the world, only in the IBM business intelligence industry spends 6 Billion of dollars for years since 5 years. Watson is accessible through a cloud open platform called Bluemix (see fig.2), has spread to 35 countries, with 18,000 programmers who work there and 500 people making their business all over the world. Watson currently has more than 30 services, 15 underlying platform technologies, and thanks to bluemix, enables the output of its API to anyone who requests it.

Figure 2.   Watson and their services (boiler plate) inside Bluemix Paas

The big data analytics is the key of managing the learning processes [13][14]. This allows us to understand and observe the students careers during their training to provide them with assistance and suggestions about the composition of university curricula, management training curricular and extracurricular, insights and the advice to acquire specific skills. New generations of (special) applications that precisely interact with a cognitive system (and therefore are named

cognos) are starting to appear. An interesting experience of using cognitive computing in educational settings is the Watson Student Showcase [8] organized by IEEE jointly with IBM. Students were challenged in a competition where they had to develop cognitive apps, using the cognitive computing services included in Bluemix.

The cognitive computing and the internet of things (IOT) will be the real innovation of the future. This will have the disruptive implications in the way of teaching [9]. If you only think about how our laboratories will be managed and maintained, with everything on cloud and with students interacting with equipment and robots that will in turn interact with other things using simple interfaces and often through the voice. all this can be achieved by developing a new generation of cognitive app that uses the voice and therefore might call cognitive app (Cognos).

Cognos learn and communicate with other machines making it easy to manage the interaction between man and machine and from machine to machine. Consider the impact of all this on our teaching laboratories: electrical and electronic measurements, computer science, robotics etc. Labs may be automated, students will be able to use them at anytime from anywhere, without limitation of time. An approach of this type will revolutionize education, teaching and the contents. Of course cognitive computing as Watson and other can make this already possible [7].

Another interesting side is the bureaucratic management of teachers and students. For the management of students, the observation is not only limited to the period of University residence, but can be extended before (in high school) and after (in employment) in order to have a traceability of the careers of the skills and knowledge acquired. All this is achieved through the management of big data and through the management of platforms dedicated to this type of objective. Again cognitive systems [5] well compete through the analytics. Even the management of teachers is interesting through analytics, and more generally, many problems can be managed through analytics foreshadowing a smarter Universities [4].

The analytics can also be extended to networks of people who share a common goal. The *nose of knowledge* helped a lot in building a network of expertise and knowledge [7]. But this knowledge is often limited to students in a class network. This causes the potential of the group remain limited. If we were interested to boost the potential of this network group should be extended to accommodate the students' knowledge and skills from other classes, other universities geographically located distant from us. Precisely in this regard you may develop a *nose of knowledge cognos app* that extends the capabilities of another app developed to the Ohio state university.

In fact The **Ohio State University** created a web application for the Watson University Competition called the "**YouDoU**" [15] project that helps Ohio State students find activities based on their personal preferences using Watson technology. An extension of personal preferences are also represented by the skills

and knowledge acquired during their university studies that everyone wants to pool in a workgroup. This extension can be managed by ***nose of knowledge cognos app.*** Many analytics can be used. Analytics can be used to measure the reputation that each student acquires in each group, by virtue of the activity that takes place during the collaboration. This way we can dynamically reconfigure the network by combining the elements that achieve the highest score, excluding those that bear a lesser reputation of a certain threshold, etc.

Finally analytics can also be applied on Italian University's admission tests to identify the University course credits that each student must mature and then follow him in his career and give him a concrete and effective customized training support.

V. CONCLUSIONS

In this paper we presented some application scenarios of big data on e-learning. More generally, the authors think that the ability to think, learn and understand the world around them will be essential to compete in the near future both for universities and companies . All this will be possible by the use of big data. We are surrounded by big data, and it is natural to use them to do business through analytics. In Universities, it is natural to use big data to manage learning processes in an intelligent way and to help individual students to improve their skills.

From the perspective of companies, we are facing a very strong disruptive technology that will transform everything, i.e., cognitive computing. Technologies related with analytics, cloud , mobility, security are very important, but one question arise to which we have to find an answer: when will all companies be digitized enough to compete in the market and win the competition?

The point is that the process of digitization in a company is not among its primary goal, but constitutes the premises for further digital revolution: the cognitive IOT. The two drivers of this new technology will be invisible and cognitive data. Eighty percent of data is invisible and is not understandable, "obscure" so to say, because it has not a structure but is also stored somewhere. This includes video, music, news articles, research reports, social media posts, enterprise data system etc. The second driver is even stronger and is represented by cognitive that no longer requires programming.

Cognitive systems in fact require the use of spoken language and there are currently many existing applications, although at an experimental level, on sails, supply chain, research, child (with *dyno* cognos), bluematch (putting together positions, persons and curricula of each one in order to reach specific project objectives), education , cooking recipes, medicine, meteorology, etc.). The difference between the cognitive and the current systems is that systems understand, reason and learn. It 'a revolution in technology, business and teaching.

The revolution in teaching will be even stronger because it will change the paradigm, the educational materials and the laboratories, and this in turn will imply the change in the role of the teacher, that will have a co-coach role assisted by cognitive systems. The support to this new paradigm of teaching and learning will cause a circular mechanism, where more and more data will be needed. So the sooner the e-learning world will realize and implement big data inside their technologies and processes, the better for the whole e-learning movement.

REFERENCES

[1] E. AbuKhousa and Y. Atif (2014), "Big learning data analytics support for engineering career readiness," Interactive Collaborative Learning (ICL), 2014 International Conference on, Dubai, 2014, pp. 663-668.

[2] Banavar, G.S. (2015), *Watson and the era of cognitive computing,* in Proceedings of the 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom), 95-95.

[3] Byrd, G. (2016), IEEE/IBM Watson student showcase, IEEE Computer, 49 (1), 102-104.

[4] Coccoli M., Guercio A., Maresca P. and Stanganelli L. (2014) "Smarter Universities: a Vision for the Fast Changing Digital Era", in Proceedings of 20th International Conference on Distributed Multimedia Systems (DMS2014)

[5] Coccoli M., Maresca P., Stanganelli L. (2012) "Cognitive computing in Education ", Journal of e-Learning and Knowledge Society (Je-LKS),Vol. 12, n. 2, May 2016, pp. 55 – 69.

[6] Ferguson, R. and Buckingham Shum, S. (2012). Social Learning Analytics: Five Approaches. LAK 2012. 23--33. DOI: 10.1145/2330601.2330616

[7] Gašević, D., Dawson, S., & Siemens, G. (2015). Let's Not Forget: Learning Analytics are about Learning. Techtrends, 59(1), 64--71. DOI: 10.1007/s11528-014-0822-x

[8] Holtel, S. (2014), *More the end of information over ow: how IBM Watson turn upside down our view on information appliances*, in Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-I.T), 187-188.

[9] Coccoli M., Maresca P., Stanganelli L., and Guercio A. (2015) "Teaching Computer Programming in a Platform as a Service Environment", in Proceedings of 21th International Conference on Distributed Multimedia Systems (DMS2015)

[10] Riffai M. M. M. A., Duncan P., Edgar D. and Al-Bulushi A. H., "The potential for big data to enhance the higher education sector in Oman," 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 2016, pp.1-6. doi: 10.1109/ICBDSC.2016.7460346

[11] Siemens, G. & Long P. (2011) Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review 46 (5), 2011, pp. 30-38.

[12] Sundararajan, S.C. and Nitta, S. (2015), Designing engaging intelligent tutoring systems in an age of cognitive computing, IBM Journal of Research and Development, 59(6), pp. 1-9.

[13] Vozniuk, A., Govaerts, S., & Gillet, D. 2013. Towards Portable Learning Analytics Dashboards. ICALT 2013, IEEE, 412--416. DOI: 10.1109/icalt.2013.126

[14] West, D. 2012. Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. Governance Studies at Brookings, pp. 1-10.

[15] Youdou, https://www.youtube.com/watch?v=LiOI07dQVTI

[16] Colazzo, L.; Molinari, A.; Villa, N.; , "Collaboration vs. Participation: The Role of Virtual Communities in a Web 2.0 World," Education Technology and Computer, 2009. ICETC '09. International Conference on , vol., no., pp.321-325, 17-20 April 2009