Learnometrics: Metrics for Learning Objects

Xavier Ochoa
Escuela Superior Politécnica del Litoral
Via Perimetral Km. 30.5
Guayaquil, Ecuador
xavier@cti.espol.edu.ec

ABSTRACT

The field of Technology Enhanced Learning (TEL) in general, has the potential to solve one of the most important challenges of our time: enable everyone to learn anything, anytime, anywhere. However, if we look back at more than 50 years of research in TEL, it is not clear where we are in terms of reaching our goal and whether we are, indeed, moving forward. The pace at which technology and new ideas evolve have created a rapid, even exponential, rate of change. This rapid change, together with the natural difficulty to measure the impact of technology in something as complex as learning, has lead to a field with abundance of new, good ideas and scarcity of evaluation studies. This lack of evaluation has resulted into the duplication of efforts and a sense of no "ground truth" or "basic theory' of TEL. This article is an attempt to stop, look back and measure, if not the impact, at least the status of a small fraction of TEL, Learning Object Technologies, in the real world. The measured apparent inexistence of the reuse paradox, the two phase linear growth of repositories or the ineffective metadata quality assessment of humans are clear reminders that even bright theoretical discussions do not compensate the lack of experimentation and measurement. Both theoretical and empirical studies should go hand in hand in order to advance the status of the field. This article is an invitation to other researchers in the field to apply Informetric techniques to measure, understand and apply in their tools the vast amount of information generated by the usage of Technology Enhanced Learning systems.

Categories and Subject Descriptors

K.3 [**Computers and Education**]: General

General Terms

Learnometrics, Learning Analytics, Metrics

Keywords

Metrics, Learning Object, Repositories, Reuse

1. INTRODUCTION

This dissertation presents the measurement of several characteristics related to a learning object and the different processes that take place during its life cycle. We have called this initiative "Metrics for Learning Objects" or "Learnometrics" for short. This denomination was chosen to reflect the similarity of this study goal and methodology with the Informetric fields. For example, Bibliometrics, which is the scientific field focused on the measurement and analysis of texts and information [5], Scientometrics, which measures the scientific process [17] and Webometrics, that analyzes the behavior of the World Wide Web and the Internet [1].

Informetrics is focused on measuring and understanding processes that create, publish, consume or adapt information. Moreover, it is common that after the process has been analyzed, useful metrics are developed to summarize characteristics of the process and then used to create tools that can have a practical application to improve the studied or a related process. Scientometrics, for example, has studied the scientific publication and citation processes. Extensive publication and citation data has been quantitatively analyzed. From these analyzes, it has been found that the number of publications per author and the number of citations per journal follows the Lotka law [21]. Based on these findings, several scientists have suggested models that explain the publication and citation process. "Success breeds success" [11] and "Cumulative advantage" [32] are two ways to express that the probability to publish a new scientific article or receive a new citation is proportional to how many articles the author has published before or how many citations the journal has already received. Practical metrics that have been extracted from these analyses are the Journal Impact Factor [14] and the h-index [16]. Those metrics serve to summarize the scientific impact that a journal or a scientist has in a particular field. Moreover, these metrics, while not perfect [18], are often used as selection criteria in other scientific processes, such as when selecting a journal to publish research or selecting the most talented scientist to fulfill an academic position.

This Learnometric study follows a similar pattern. It quantitatively analyzes data from processes that take place during different points of the Learning Object life cycle. Based on that analysis, initial models are proposed to explain the observed results. The study also proposes small calculations (metrics) that can covert the data available about the learning objects into information that can be used to improve the effectiveness or usefulness of existing Learning Object end-user tools.

2. UNDERSTANDING THE PUBLICATION OF LEARNING OBJECTS

The first step to understand the Learning Object Economy is to measure and understand how learning objects are offered or published. A literature review on the topic is very discouraging. The only serious work that tries to quantitatively measure the publication of learning objects is [22], which is, however, very superficial on its quantitative side and draws no conclusions from the results. This lack of research leads to an almost unexplored field with even the most basic questions unanswered.

This section will quantitatively analyze and compare different types of publication venues for learning objects . These types include Learning Object Repositories (LORP), Learning Object Referatories (LORF), Open Courseware Initiatives (OCW), Learning Management Systems (LMS). To provide some type of comparison and because their content can also be used for educational purposes, Institutional Repositories (IR) are also included in the study. For simplicity, we will refer to all these systems as “repositories”. The following subsections provide the analysis of data collected from this repositores to provide answer to several basic questions.

2.1 What is the typical size of a repository?

This analysis measured the size distribution of different types of repositories. This study included data from 24 LORPs, 15 LORFs, 34 OCWs, 2500 LMSs and 772 IRs. More details about the repository selection and data collection could be found in [31]. Figure 1 shows the range of the obtained sizes (y-axis in log scale) for each type of repository. In general, individual learning object repositories seems to vary from hundreds to million of objects. Their average size depends of the type of repository. LORPs can be considered to have few thousand of objects. LORFs are in the order of the tens of thousands. However, those numbers are small compared with multi-institutional IRs that can count hundreds of thousands and even millions of objects. OCWs and LMSs can have from hundreds to thousand of courses with a total of thousands or ten of thousands of individual resoures. However, the answer to this question is not that simple. The size is not Normally distributed, meaning that the average value cannot be used to gain understanding of the whole population. It is not strange to find repositories several orders of magnitude bigger or smaller than the average. The distribution of learning objects among repositories seems to follow a Lotka or Power Law distribution with an exponent of 1.75. The main implication of this finding is that most of the content is stored in few big repositories, with a long, but not significant tail. Administrators of a big repository would want to federate [34]] their searches with other big repositories in order to gain access to a big proportion of the available content. On the other hand, it makes more sense for small repositories to publish their metadata [37] for a big repository to harvest it in exchange for the access to their federated search. It seems, through an initial reading of this finding, that a two (or three) tiered approach mixing federation and metadata harvesting is the most efficient way to make most of the content available to the wider audience possible using the current infrastructure.

Figure 1: Size range of different type of repositories

Another important implications of thes findings is that if OCWs and LMSs are decomposed and converted into repositories, they can be considered very large LORPs. The fact that LMSs are a widely deploy technology [15] and that these systems are not accessible for external visitors make us think of the learning objects present in LORs as just the “tip of the iceberg’. The bigger part of learning resources is hidden behind login pages. This finding validates the effort of the OCW Consortium and OER Commons [19]. If we want to create a really functioning Learning Object Economy, we must start opening the door of our LMSs.

2.2 How do repositories grow over time?

To measure the growth in the number of objects, 15 repositories of different type were studied. They were selected based on how representative they are for their respective type in terms of size and period of existence. The first variable analyzed was the average growth rate (AGR), measured in objects inserted per day. It is interesting to compare the AGR of different types of repositories. LORPs, for example, grow with a rate of 1 or 2 objects per day. LORFs goes from 4 to 20 objects per day. OCWs and LMSs grow faster than LORPs, with an unexpectedly high value of circa 1 course published per day (in average 20 objects). IRs depending on their size could go from few objects to hundreds of objects per day, depending on their size. The actual growth function for most respositories is linear (bi-phase linear).

This is a discouraging finding. Even popular and currently active repositories grow linearly. Even if we add them all together, we will still have a faster linear, but no exponential. The main reason for this behavior is the contributor desertion. Even if the repository is able to attract contributors, it is not able to retain them long enough. The value proposition, that is the way how the contributor benefits from contributing to the repository, is still an unsolved issue in most repositories.

One anomaly in this study was the LORP Connexions. It grows at what seemed to be an exponential rate. Figure 2 shows this difference. Further investigagtion in [27] revealed that the social features present in Connexions that enable the formation of communities around the materials are the reason behind Connexions sucess. Based on these results, it seems that the use of social engagement tools should be part of any new repository design.

2.3 How many learning objects does a contributor publish on average?

To understand contributor behavior, full publication data from three LORPs (Ariadne, Connexions and Maricopa), one LORF (Merlot), one OCW site (MIT OCW), one LMS (SIDWeb) and three IRs (Queensland, MIT and Georgia Tech) was obtained. Each learning object was assigned according to the data to one contributor. If more than one contributor was listed, we counted the first author only. From the result of the distribution fitting, it is clear that the number of objects published per each contributor varies according to the type of repository. All LORPs and LORFs follow a Lotka distribution with exponential cut-off. Even high producing individual start loosing interest after publishing many objects. Maybe one of the reasons behind this distribution is the lack of some type of incentive mechanism [6]. OCW MIT and SIDWeb present a Weibull distribution. The finding of a weibull distribution means that for OCWs and LMSs there is an increased probability to produce a certain amount of objects. This can be seen as the strong concavity in the curve compared with the flat Lotka. The mechanism behind this distribution is that there is an interest to produce courses with a given amount of learning objects (maybe 1 object per session). The tail of the IRs are fitted by the pure Lotka distribution. The head of the distribution, users that have published 1 or 2 objects, have a disproportionately high value that cannot be fit by any of the tried distributions. This result suggests that the publication of documents in IRs have a different mechanism than the publication of learning objects in LORs, and maybe what we are measuring in the IRs tail is a by-product of the scientific publication process. These distributions could be seen in Figure 3.

Based on the finding of these heavy tailed distribution, it can be concluded that "there is not such thing as an average user" [28]. The best way to describe the production of different contributors is to cluster them in "classes" similar to socioeconomic strata. If we adopt this approach we gain a new way to look at our results. In LORP and LORF, the repository is dominated by the higher-class. Most of the material is created by a few hyper-productive contributors. the 10% of the users could easily have produced more than half of the content of the repository. In the case of OCWs and LMS, the Weibull distribution determines that the middle-class is the real motor of the repository. The low- and high-class are comparatively small. Finally, University IRs, with Lotka with high *alpha* are dominated by the lower-class as more than 98% of the population produces just one object.

From a deper analysis on publishing rate and lifetime [31], it can be concluded that these different distributions are not caused by an inherent difference in the talent or capacity among the different communities, but by the difference in contributor engagement with the repository. It seems that the distribution of lifetime, the time that the contributor remains active, is different for this three observed repository types. In LORP and LORF, there is some time of novelty engagement that keep the contributor active at the beginning, but the chances of ceasing publication increases as more time is spent in the repository. For OCWs and

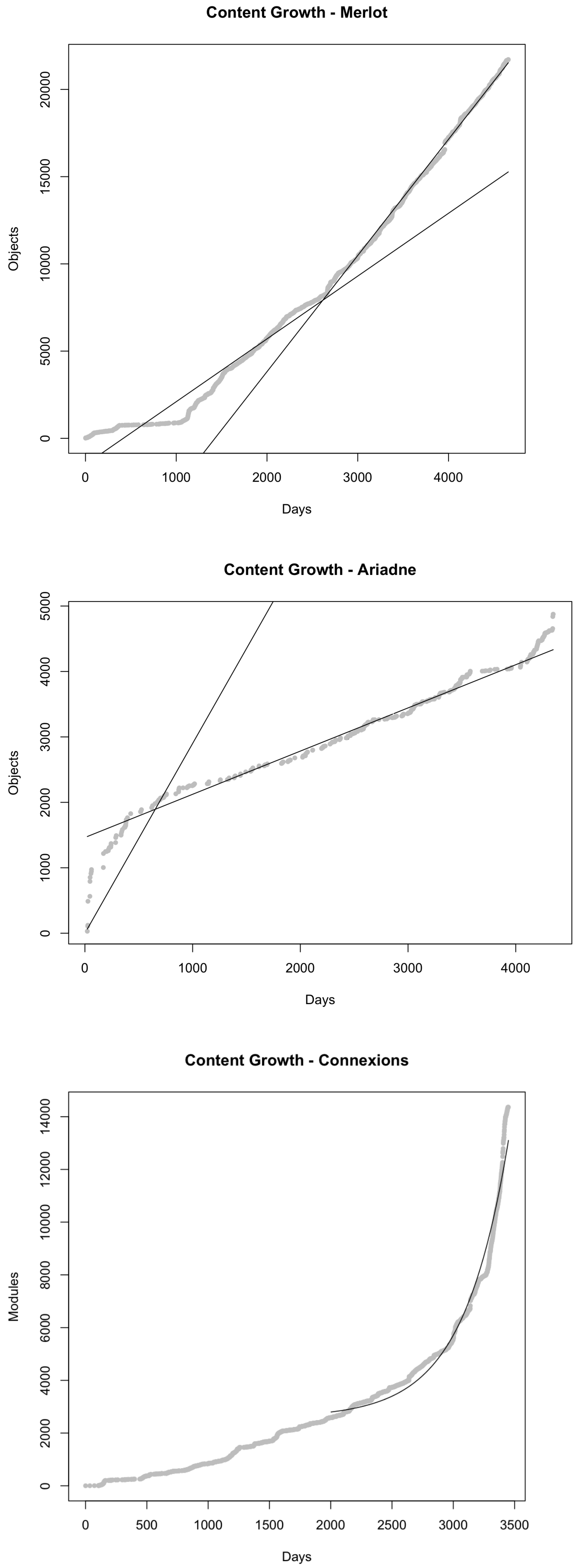

Figure 2: Comparation of Content Growth Function for Connexions, Ariadne and Merlot

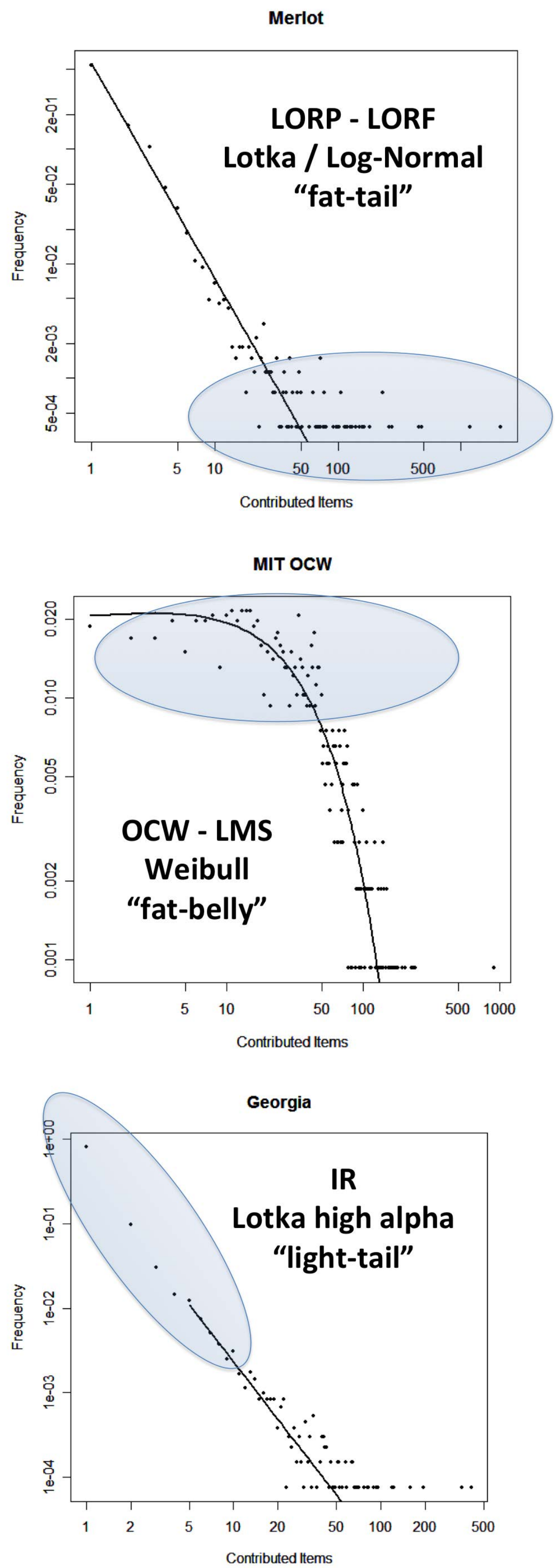

Figure 3: Distribution of Contribution for the different type of repositories

LMSs, there is a goal-oriented engagement that keeps the contributor productive until her task is finished (course is fully published). In the case of IRs, there is no engagement at all. The norm is just discrete contributions. Changes on the type of engagement should have an effect not only in the distribution of publications among users, but also in the growth and size of the repository.

3. UNDERSTANDING THE REUSE OF LEARNING OBJECTS

Although reuse is the reason why much of Learning Object Technologies exist, little is quantitatively known about the Reuse process. Beside small scale experiment in artificial settings [33] [12] [38], there is practically no empirical data on how different factors affect the reusability of learning objects. Again, with an almost unexplored field, this article proposes and aims to solve the some basic questions.

3.1 What percentage of learning objects is reused?

To perform a quantitative analysis of the reuse of learning objects, this study uses empirical data collected from three different openly available sources. The sources were chosen to represent different reuse contexts and different object granularity.

Small Granularity: Slide Presentation Components. A group of 825 slide presentations obtained from the ARIADNE repository [10] were decomposed and checked for reuse using the ALOCOM framework [39]. From the decomposition of the slides 47,377 unique components were obtained. A component is considered reused if it is present in more than one slide.

Medium Granularity: Learning Modules. The 5255 learning objects available at Connexions [3] at the time of data collection were downloaded. Some of these objects belong to collections, a grouping of a similar granularity as a course. 317 collections are available at Connexions. A module is considered reused if it is used in more than one collection.

Large Granularity: Courses. The 19 engineering curricula offered by ESPOL, a technical University at Ecuador, reuse basic and intermediate courses. When a new curriculum is created, existing courses, such as Calculus and Physics, are reused. On the other hand, more advanced courses, for example Power Lines in the case of Power Engineering, are created and only used in the specific curriculum. Based on the published information, the 463 different courses were obtained. A course is considered reused if it is mandatory in more than one curriculum.

The results of the quantitative analysis (Table 1) seems to indicate that in common settings, the amount of learning objects reused is around 20%. While relatively low, this result is very encouraging for Learning Object supporters. It indicates that even without support or the proper facilities, users do reuse a significant amount of learning materials. The multiplicative model also implicates that improving even one of the steps in the reuse chain, the others remaining equal, would improve the probability of reuse and, therefore, the amount of objects being reused. As mentioned above, Verbert and Duval, in [38], empirically found that facilitating one of the steps, in this particular finding slide components, leads to a significant increase in the amount of reuse.

3.2 Does the granularity of a learning object affect its probability of reuse?

Table 1: Percentage of reuse

Data Set	Objects	% Reuse
Small Granularity		
Components in Slides (ALOCOM)	47,377	11.5%
Images (Wikipedia)	1,237,105	24.6%
Medium Granularity		
Modules in Courses (Connexions)	5,255	22.6%
Soft. Libraries (Freshmeat)	2,643	20.4%
Large Granularity		
Courses in Curricula (ESPOL)	463	19.9%
Web APIs (P.Web)	670	32.2%

The theory of Learning Objects affirms that higher granularity leads to lower reusability [40] . Results from the previous study, however contradict this affirmation. The percentage of object reuse was similar regardless of the granularity of the object. Courses were even reused more often than slide components. Merging the theory with the empirical finding leads to a new interpretation of the role of granularity in the reuse of learning objects. This new interpretation involves also the granularity of the context of reuse as the determining factor. Objects that have a granularity immediately lower than the object being built are easier to reuse than objects with a much lower or higher granularity. For example, when building a course, it is easier to reuse whole lessons than reusing complete courses or individual images. Also, when building a curriculum, it is easier to reuse complete courses than to reuse another complete curriculum or individual lessons. Empirical support for this new interpretation can be found in [38]. It was found that when building a slide presentation, the most reused component was by far individual slides. The reuse of text fragments and individual images represent just the 26% of the total reuse.

3.3 Is there a relation between the popularity of an object and its reuse?

The objective of this analysis is to establish if the actual reuse of a learning object is linked to its relative popularity within the collection or repository. To perform this analysis, the Connexions and Freshmeat data sets were enriched with information about the number of times that the objects have been accessed. The popularity data was obtained from Web scraping These data sets were selected for this analysis because they were the only ones with access information and have similar granularity.

The analysis consisted in obtaining the Kendall's tau correlation coefficient between the rank of the object in the reuse and popularity scales. Pearson's coefficient is not used because there is no guaranty that the values come from a bi-variate normal distribution. Also, scatter plots were created to visually analyze the relation between popularity and reuse. Figure 4 presents the data for Connexions. The correlation coefficient tau for the Connexions set was -0.02 (0.05 significant). This value means that there is absolutely no correlation between the popularity of the object and the times that it has been reused. For example, the most visited object has only be reused in three collections, while the

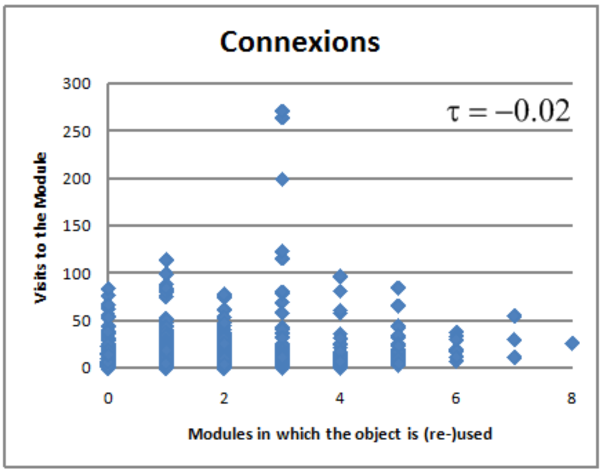

Figure 4: Scatter plots of the Reuse vs. Popularity in Connexions

most reused object (8 times) has only received 25 visits. On the other hand, the Freshmeat set obtained a tau of 0.33 (0.01 significant). This result suggests that in the case of software libraries the popularity is slightly linked with the reuse. However, there are cases that have a large popularity but have a low track of reuse. For example, the DeCSS library [13], normally used to break DVD encryption, has a large popularity (circa 180.000 visits) but is only used in a small set of specialized DVD players for Linux (8 projects). These results suggest that the popularity of an object cannot always be used as a proxy for its reuse. A more counter-intuitive finding that can be obtained from this result is that a high level of reuse does not imply a high popularity. It would be usually expected that an object reused in several contexts is more findable and, therefore, more visited. The measurement indicates that it is not the case.

3.4 What is the distribution of reuse among learning objects?

To gain more insight in the reuse process, the distribution of reuse among different objects was analyzed. The first step in this analysis was to obtain the total number of reuses for each object. Several distributions were fitted to the data to obtain the best fit. For all the data sets, the Log-normal distribution provided the best fit. As a visual aid, Figure 5 presents the size-frequency plot of the data [26].

The main implication of the finding of a Log-normal distribution is that the "Long Tail" effect [2] applies to reuse. Few objects are reused heavily while most of the reused objects are reused just once. However, the volume of reuse in the tail is at least relatively as important as the volume of reuse in the head. According to this result, federating repositories in order to provide a wider selection of objects is a good strategy to foster reuse. Objects present in small repositories have a high probability of being reuse at least once if they are exposed to a wider universe of users.

4. METRICS FOR LEARNING OBJECTS

The main use that we can give to the information extracted from the analysis of the data created at the different processes of the Learning Object Economy is the creation of

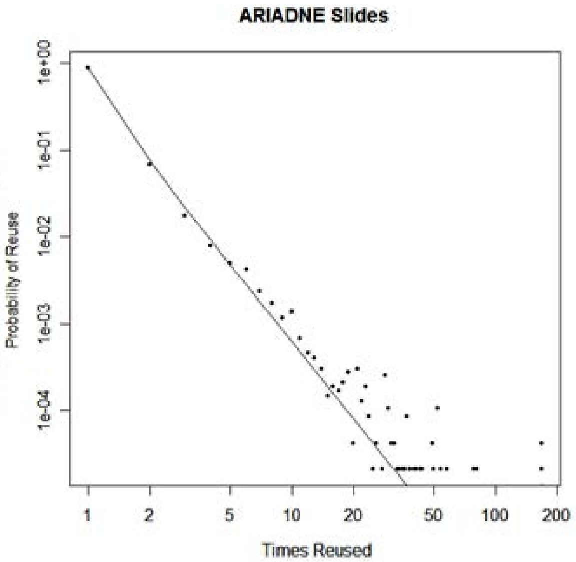

Figure 5: Size-Frequency graphs of the reuse in ARIADNE Slides and the best fitting Log-Normal distribution (line)

Figure 6: Visualization of the Textual Information Content of the ARIADNE Repository. Red (dark) boxes indicate authors that produce low quality descriptions.

metrics to improve the tools used in those processes. This article will discuss two specific examples of these metrcis: 1) to estimate the quality of the learning object metadata and 2) to establish the relevance of learning objects for a given user and situation.

4.1 Quality Control for the Labelling Process

The quality of metadata on learning objects stored in a LOR is an important issue for LOR operation [4] and interoperability [20]. Due to its importance, metadata quality assurance has always been an integral part of resource cataloging [36]. Nonetheless, most LOR implementations have taken a relaxed approach to metadata quality assurance. As repositories grow and federate, quality issues become more apparent. The traditional solution for quality assurance, manually reviewing a statistically significant sample of metadata against a predefined set of quality parameters, similar to sampling techniques used for quality assurance of library cataloguing [7], fails to scale to increasing amounts of learning objects being indexed manually or automatically. Some sort of automatical quality assurance mechanism should be created to cope with this problem.

In [30], the author propose and evaluate a set of metrics to automatically measure the quality of the learning object metadata instances. The main conclusions of this work are that some metrics correlate well with human reviews, specially the Textual Information Content (Qtinfo). Also the metrics, when combined serves as a low quality metadata filter. Figure 6 present an application to evaluate the metadata quality of whole repositories based on the proposed metrics.

4.2 Relevance Ranking to Improve the Selection Process

In the early stages of the Learning Object Economy, LORs where isolated and only contained a small number of learning objects [25]. The search facility usually provided users with an electronic form where they could select the values for their desired learning object. The search engine then compared the values entered in the query with the values stored in the metadata of all objects and returned those which complied with those criteria. While initially this approach seems appropriate to find relevant learning objects, experience shows that it presents several problems, such as high cognitive load [23], mismatch between indexers and searchers [24], and low recall [35]. Given these problems with the metadata based search, most repositories provided a "Simple Search" approach, based on the success of text based retrieval exemplified by Web Search engines [8]. In this approach, users only need to express their information needs in the form of keywords or query terms. This approach seemed to solve the problems of metadata based search for small repositories. However, working with small, isolated repositories also meant that an important percentage of users did not find what they were looking for because no relevant object was present in the repository [23]. If this technique is applied to large repositories, or to federated collections of repositories, the user is no longer able to review several pages of results in order to select the relevant objects. While doing a stricter filtering of results (increasing precision at expense of recall) could solve the oversupply problem, it could also lead again to the initial problem of scarcity. A proven solution for this problem is ranking or ordering the result list based on its relevance. In this way, it does not matter how long the list is, because the most relevant results will be at the top and the user can manually review them.

In a previous work [29], the author describes a set of relevance ranking metrics for learning objects. These metrics try to implement the theoretical LearnRank [9]. This work found that the information about the usage of the learning objects, as well as the context where this use took place, can

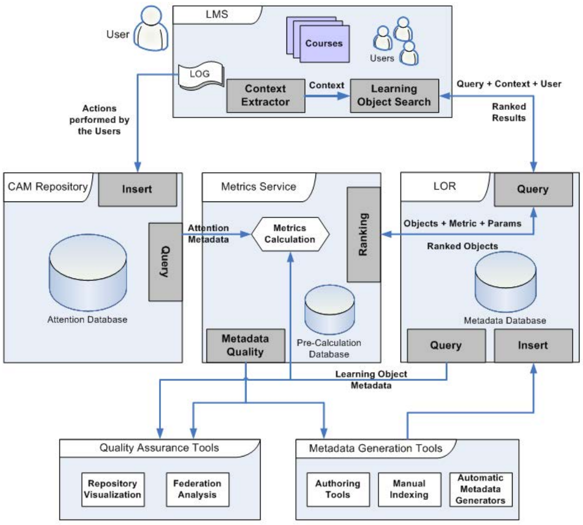

Figure 7: Architecture for Metrics Services

be converted into a set of automatically calculable metrics to establish the relevance of a learning objects for a given user in a given situation. The evaluation of the metrics show that these metrics outperformed the ranking based on pure text-based approach. Figure 7 presents an architecture discussed in [29] to implement these metrics in real systems.

5. NOT CONCLUSIONS BUT FURTHER RESEARCH

As a first exploration of Learnometrics, this article, and its cited studies, raises more questions than it answers. Ample opportunities for further research are provided as the field of Learnometrics unfolds. The following is a list of what the author consider are the most interesting and urgent research questions seeking for answers and explanations.

- What is the measurable effect that opennes have in the Learning Object Economy
- How to integrate LMSs into the Learning Object Economy
- How to reformulate the Paradox of Reuse to consider more varibles apart from granularity
- Establishing a common data set to experiment with metrics and their usefulness

Answering these questions through quantitative analyses will increase our understanding of how the Learning Objects Economy works. This understanding can help us to create the right environment for this economy to flourish and provide its predicted benefits. The main task left for further work is to execute large empirical studies with full implementations of the metrics in real environments. Once there is enough data collected, the user interaction with the system and the progress of the different metrics could be analyzed to shed light on these questions. We also hope that other researchers start proposing improvements to these initial approaches.

6. REFERENCES

[1] T. C. Almind and P. Ingwersen. Informetric analyses on the world wide web: methodological approaches to ŚwebometricsŠ. *Journal of Documentation*, 53(4):404–426, 1997.

[2] C. Anderson. *The long tail*. Hyperion, 2006.

[3] R. G. Baraniuk. *Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge*, chapter Challenges and Opportunities for the Open Education Movement: A Connexions Case Study, pages 116–132. MIT Press, 2007.

[4] J. Barton, S. Currier, and J. M. N. Hey. Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. In S. Sutton, J. Greenberg, and J. Tennis, editors, *Proceedings 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications*, pages 39–48, Seattle, Washington, 2003.

[5] R. Broadus. Toward a definition of ŞbibliometricsŤ. *Scientometrics*, 12(5):373–379, 1987.

[6] L. Campbell. *Reusing Online Resources: A Sustainable Approach to E-Learning*, chapter Engaging with the learning object economy, pages 35–45. Kogan Page Ltd, 2003.

[7] A. Chapman and O. Massey. A catalogue quality audit tool. *Library Management*, 23(6-7):314–324, 2002.

[8] H. Chu and M. Rosenthal. Search engines for the world wide web: A comparative study and evaluation methodology. In S. Hardin, editor, *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, volume 33, pages 127–135, Baltimore, MD, 1996. Softbound.

[9] E. Duval. *Policy and Innovation in Education - Quality Criteria*, chapter LearnRank: the Real Quality Measure for Learning Materials, pages 457–463. European Schoolnet, 2005.

[10] E. Duval, K. Warkentyne, F. Haenni, E. Forte, K. Cardinaels, B. Verhoeven, R. Van Durm, K. Hendrikx, M. Forte, N. Ebel, et al. The ariadne knowledge pool system. *Communications of the ACM*, 44(5):72–78, 2001.

[11] L. Egghe and R. Rousseau. Generalized success-breeds-success principle leading to time-dependent informetric distributions. *Journal of the American Society for Information Science*, 46(6):426–445, 1995.

[12] K. Elliott and K. Sweeney. Quantifying the reuse of learning objects. *Australasian Journal of Educational Technology*, 24(2):137–142, 2008.

[13] K. Eschenfelder and A. Desai. Software as Protest: The Unexpected Resiliency of US-Based DeCSS Posting and Linking. *The Information Society*, 20(2):101–116, 2004.

[14] E. Garfield. The impact factor. *Current Contents*, 25(20):3–7, 1994.

[15] C. Harrington, S. Gordon, and T. Schibik. Course management system utilization and implications for practice: A national survey of department chairpersons. *Online Journal of Distance Learning Administration*, 7(4):13, 2004.

[16] J. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.

[17] W. Hood and C. Wilson. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2):291–314, 2001.

[18] P. Jacsó. A deficiency in the algorithm for calculating the impact factor of scholarly journals: The journal impact factor. *Cortex*, 37(4):590–594, 2001.

[19] A. Joyce. OECD Study of OER: Forum Report. Technical report, UNESCO, 2007.

[20] X. Liu, K. Maly, M. Zubair, and M. L. Nelson. Arc - an oai service provider for digital library federation. *D-Lib Magazine*, 7(4):12, 2001.

[21] A. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12):317–323, 1926.

[22] R. McGreal. A typology of learning object repositories. [pre-print]. Retrieved December 19, 2007 from http://hdl.handle.net/2149/1078, 2007.

[23] J. Najjar, J. Klerkx, R. Vuorikari, and E. Duval. Finding appropriate learning objects: An empirical evaluation. In A. Rauber, S. Christodoulakis, and A. M. Tjoa, editors, *Proceedings of : 9th European Conference on Research and Advanced Technology for Digital Libraries. ECDL 2005*, volume 3652 of *Lecture Notes in Computer Science*, pages 323–335, Vienna, Austria, 2005. Springer Verlag.

[24] J. Najjar, S. Ternier, and E. Duval. User behavior in learning objects repositories: An empirical analysis. In L. C. . C. McLoughlin, editor, *Proceedings of the ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 4373–4378, Chesapeake, VA, 2004. AACE.

[25] F. Neven and E. Duval. Reusable learning objects: a survey of lom-based repositories. In M. Muhlhauser, K. Ross, and N. Dimitrova, editors, *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 291–294, New York, NY, 2002. ACM Press.

[26] M. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[27] X. Ochoa. Connexions: a social and successful anomaly among learning object repositories. *Journal of Emerging Technologies in Web Intelligence*, 2(1):11–22, 2010.

[28] X. Ochoa and E. Duval. Quantitative analysis of user-generated content on the web. In D. De Roure and W. Hall, editors, *Proceedings of the First International Workshop on Understanding Web Evolution (WebEvolve2008)*, pages 19–26, Beijing, China, 2008. Web Science Research Initiative. ISBN: 978 085432885 7.

[29] X. Ochoa and E. Duval. Relevance ranking metrics for learning objects. *IEEE Transaction on Learning Technologies*, 1(1):15, 2008. in Press.

[30] X. Ochoa and E. Duval. Automatic evaluation of metadata quality in digital libraries. *International Journal of Digital Libraries*, 10(2):67–91, 2009.

[31] X. Ochoa and E. Duval. Quantitative analysis of learning object repositories. *IEEE Transactions on Learning Technologies*, 2(3):226–238, 2009.

[32] D. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5-6):292–306, 1976.

[33] V. Schoner, D. Buzza, K. Harrigan, and K. Strampel. Learning objects in use:ŚliteŠassessment for field studies. *J. Online Learning Teaching*, 1(1):18, 2005.

[34] B. Simon, D. Massart, F. van Assche, S. Ternier, E. Duval, S. Brantner, D. Olmedilla, and Z. Miklos. A simple query interface for interoperable learning repositories. In D. Olmedilla, N. Saito, and B. Simon, editors, *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems*, pages 11–18, Chiba, Japan, 2005. CEUR.

[35] L. Sokvitne. An evaluation of the effectiveness of current dublin core metadata for retrieval. In *Proceedings of VALA (Libraries, Technology and the Future) Biennial Conference*, page 15, Victoria, Australia, 2000. Victorian Association for Library Automation Inc.

[36] S. E. Thomas. Quality in bibliographic control. *Library Trends*, 44(3):491–505, 1996.

[37] H. Van de Sompel, M. Nelson, C. Lagoze, and S. Warner. Resource Harvesting within the OAI-PMH Framework. *D-Lib Magazine*, 10(12):1082–9873, 2004.

[38] K. Verbert and E. Duval. Evaluating the ALOCOM Approach for Scalable Content Repurposing. In E. Duval, R. Klamma, and M. Wolpers, editors, *Creating New Learning Experiences on a Global Scale: Proceedings of the Second European Conference on Technology Enhanced Learning*, volume 4753, pages 364–377, Crete, Greece, 2007. Springer.

[39] K. Verbert, E. Duval, M. Meire, J. Jovanovic, and D. Gasevic. Ontology-Based Learning Content Repurposing: The ALOCoM Framework. *International Journal on E-Learning*, 5(1):67–74, 2006.

[40] D. Wiley, S. Waters, D. Dawson, B. Lambert, M. Barclay, D. Wade, and L. Nelson. Overcoming the limitations of learning objects. *Journal of Educational Multimedia and Hypermedia*, 13(4):507–521, 2004.