Quantitative Analysis of Learning Object Repositories

Xavier Ochoa and Erik Duval, *Member*, *IEEE*

Abstract—This paper conducts the first detailed quantitative study of the process of publication of learning objects in repositories. This process has been often discussed theoretically, but never empirically evaluated. Several question related to basic characteristics of the publication process are raised at the beginning of the paper and answered through quantitative analysis. To provide a wide view of the publication process, this paper analyzes four types of repositories: Learning Object Repositories, Learning Object Referatories, Open Courseware Initiatives, and Learning Management Systems. For comparison, Institutional Repositories are also analyzed. Three repository characteristics are measured: size, growth, and contributor base. The main findings are that the amount of learning objects is distributed among repositories according to a power law, the repositories mostly grow linearly, and the amount of learning objects published by each contributor follows heavy-tailed distributions. The paper finally discusses the implications that this findings could have in the design and operation of Learning Object Repositories.

Index Terms—Learning objects, publication, repositories, LOR, OCW, LMS.

✦

---

1 INTRODUCTION

LEARNING Object publication can be defined as the act of making a learning object available to a certain community. Collis and Strijker call this process "Offering" in their Learning Object Life-cycle model [1]. The publication process can take several forms. A professor can publish lectures notes for students in a Learning Management System (LMS). The same professor can decide to share objects with a broader community and publish them in a Learning Object Repository (LOR), such as ARIADNE [2] or Connexions [3]. The University where this professor works can decide to start an Open Courseware (OCW) initiative [4] and put the learning material of its courses freely available on the Web. Moreover, material already available online can be discovered and republished for other communities. For example, a student that found an interesting Web site to learn about basic Physics could publish a link to that Web site on a Learning Object Referatory (LORF), such as Merlot [5] or SMETE [6]. In all its different forms, Learning Object publication is the most important enabler of the Learning Object Economy [7], because making the objects available is the first step to fuel the "share, reuse, improve, and share again" philosophy behind this economy.

The publication of learning objects has been an important research issue since the definition of the field 15 years ago. These efforts can be summarized into three different research lines: Publishing Infrastructure [8], [2], Interoperability [9],

---

- X. Ochoa is with the Centro de Tecnologías de Información, Escuela Superior Politécnica del Litoral, Campus Gustavo Galindo, Via Perimetral Km. 30.5, Apartado 09-01-5863, Guayaquil, Ecuador.
  E-mail: xavier@cti.espol.edu.ec.
- E. Duval is with the Department Computerwetenschappen, Katholieke Universiteit Leuven, Celestijnenlaan 200 A, B-3001 Leuven, Belgium.
  E-mail: erik.duval@cs.kuleuven.be.

and Copyright and DRM [10], [11] . However, one area of research that is practically unexplored is the study of the actual process and results of learning object publication. The research on technical and legal aspects lays the ground on which publication can take place. However, it does not provide any information about simple questions, such as how many learning objects are actually published, how they are distributed among different repositories, or how repositories grow. Moreover, answers to these questions are not only relevant to measure the progress of the Learning Object Economy, but also to provide information on which decisions about architecture, interoperability strategies, and planning for growth should be based.

The most prominent attempts to characterize learning object repositories and measure their characteristics are made by McGreal in [12]. He provides a comprehensive survey of existing LORs and classifies them in various typologies. Unfortunately, his analysis is mostly qualitative and cannot be used to answer the questions mentioned above. Other relevant studies are [13], [14], [15], [16], and [17]. In these studies, different LORs are also qualitatively compared, mainly by general characteristics as metadata standard used, language, end users, quality control, etc. In contrast with these studies, this paper will quantitatively analyze and compare different types of publication venues for learning objects. These types include Learning Object Repositories (LORPs), LORFs, OCW Initiatives, and LMSs. To provide some type of comparison and because their content can also be used for educational purposes, Institutional Repositories (IRs) are also included in the studies. For simplicity, during this paper, we will refer to all these systems as "repositories." The main goal of this paper is to provide empirical answers to the following questions:

- What is the typical size of a repository? Is it related to its type?
- How do repositories grow over time?
- What is the typical number of contributors a repository has? Is it related to its type?

- How does the number of contributors grow over time?
- How many learning objects does a contributor publish on average?

To answer these questions, data from different repositories are collected and analyzed. These answers will help us to gain insight into the actual process of learning object publication. Understanding how supply works in the Learning Object Economy will help the administrator of the repositories to design and plan the technological infrastructure needed to receive, store, and share the published material. Policy-Makers can also use these answers to evaluate which are the best approaches to encourage Contributors to publish their materials.

This paper is structured as follows: Section 2 presents an analysis of the size distribution of different repositories. Section 3 analyzes the growth rate in objects, as well as contributors. Section 4 studies the distribution of contribution, publishing rate, and engagement time. Section 5 discusses the implication of the findings and answers the research questions. The paper finalizes with Conclusions and Further Work.

2 SIZE ANALYSIS

In this section, we will analyze the size of different repositories. We define size as the number of objects present in the repository. We compare the number of objects between repositories of the same type. We start with the study of 24 LORPs and 15 LORFs. These LORs were selected from the list compiled in [12]. This list is biased toward repositories that content materials in English. To avoid an unfair size comparison between repositories, only LORs that are not the result of the federation of other repositories that are publicly available and contain or link to learning objects of small and intermediate granularity (raw material or lessons) were analyzed. While McGreal already reported an estimate of the size, we measured each LOR through direct observation on 3 and 4 November 2007.

The 24 LORPs have in total circa 100,000 learning objects, with an average size of circa 4,000 objects. A simple histogram of the data shows that the size distribution is not Normal, but highly skewed to the left. To analyze the distribution, we fit five known probabilistic distributions to the data: Lotka, Exponential, Log-Normal, Weibull, and Yule. These distributions were selected because they have high skewness to the left and are commonly present in other Information Production Processes [18]. The Maximum Likelihood Estimation (MLE) method [19] was used to obtain the distribution parameters. To find the best-fitting distribution, the Vuong test [20] is applied on the competing distributions. When the Vuong test is not statistically significant between two distributions, the distribution with less parameters is selected. This methodology is recommended by [21] to select among heavy tail models instead of the more common Least-Squares Estimation and R^2 values used for Generalized Linear Models. In the specific case of LORP, the best-fitting distribution is Exponential ($\lambda = 2.5 \times 10^{-3}$). Fig. 1a presents the empirical (points) and fitted (line) Complementary Cumulative Distribution functions (CCDFs) presented in logarithmic scales. In this graph, the X-axis represents the number of objects present in the repository. The Y-axis represents the inverse accumulated probability of the size ($P(X \geq x)$), that is, the probability
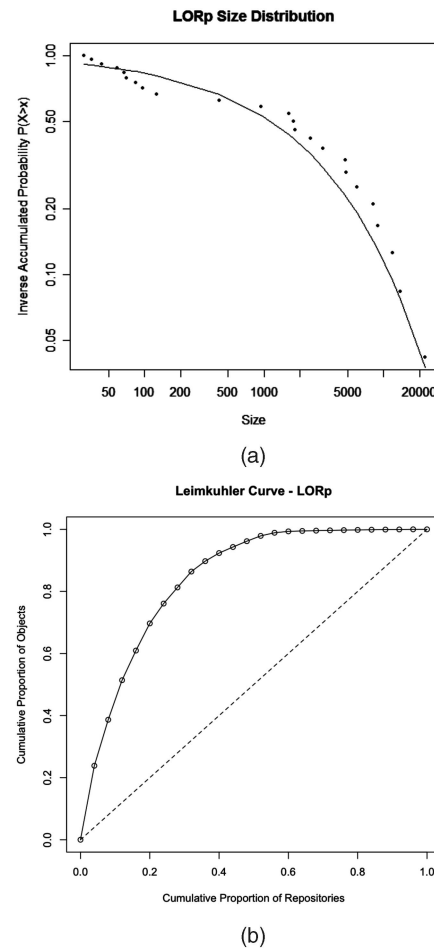

Fig. 1. Distribution and Leimkuhler curve of the size of learning object repositories.

that a repository has x or more objects. This skewed distribution of content size concentrates the majority of learning objects in few big repositories, while the rest of repositories contribute only a small percentage. Fig. 1b shows the Leimkuhler curve [22] [23]. This curve is a representation of the concentration of objects in the different repositories. The Y-axis represents the cumulative proportion of objects published in the top x proportion of repositories. For example, it is easy to see that the top 20 percent of the repositories (the biggest 5) contribute almost 70 percent of the total number of learning objects. The smaller 40 percent of the repositories combined contribute less than 3 percent of the objects.

The 15 studied LORFs offer in total circa 300,000 learning objects, with an average of circa 20,000 objects per referatory. The best-fitting distribution is Exponential ($\lambda = 5.2 \times 10^{-5}$). The biggest 20 percent (three referatories) concentrate 66 percent of the 300,000 objects. The lower half contribute only 10 percent of the total.

To study the size of OCW initiatives, we collect a list of 34 institutions providing their materials online from the Web site of the OCW Consortium.[1] The size was determined by the number of full courses offered online by each Institution. In total, 6,556 courses are available among all the studied sites, with an average of 193 courses per site. However, the

---

1. Open Courseware Consortium, http://www.ocwconsortium.org/.

TABLE 1
Summary of Number of Objects Analysis

Repository	Sampled	Average Size	Distribution	Parameters	Conc. (20%)
LORp	24	3,905	Exponential	l=2.5x10-3	70%
LORf	15	19,396	Exponential	l=5.2x10-5	66%
OCW	34	193	Lotka	a=1.61, xmin=8	90%
LMS	2,500	67	Lotka	a=1.95, xmin=70	85%
IR	775	9,820	Lotka	a=1.73, xmin=3304	90%

average is a misleading value, because the size distribution is extremely skewed. The best fitted distribution for OCW sites with more than 7 objects (tail) was Lotka with an $\alpha = 1.61$. The estimation of the start of the tail was performed minimizing the Kolmogorov-Smirnov statistic [24] for the Lotka law. This distribution leads to a very unequal concentration of courses. The top 20 percent (seven sites) of the OCW sites offer almost the 90 percent of the courses, the remaining 27 sites just account for 10 percent.

Data about the size of LMSs are not usually available online. Most LMS implementations only allow registered users to have access to their contents. In order to obtain an estimation of the size of an LMS, we use some characteristics of Moodle [25], a popular Open-Source LMS. Moodle allows guests to see the list of courses and a link to registered installations is available on the Moodle site. We obtained a random sample of 2,500 from the circa 6,000 LMS sites listed on the Moodle site as installations in the United States. This country was selected because it had the largest number of installations. Through Web scraping, we downloaded the list of courses for each one of those installations. In these 2,500 Moodle sites, 167,555 courses are offered, with an average of 67 courses per site. The distribution that best-fits the tail of the data (sites bigger than 70 courses) is Lotka with an estimated α of 1.95. The estimation of the start of the tail was determined through minimization of the Kolmogorov-Smirnoff. Again, this distribution concentrates most of the courses in just a few LMSs. For example, the top 20 percent LMSs (500 sites) offer more than 85 percent of the courses.

Finally, to establish the size distribution of IRs, we collect the list of repositories listed at the Registry of Open Access Repositories (ROARs).[2] An automated service connected to this registry regularly harvests OAI-PMH-enabled [26] IRs and provides information about their sizes. During data collection, 772 repositories with more than one object were measured. The total number of documents stored in those repositories was 7,581,175. There were, on average, 9,820 documents per repository. The tail (repositories with more than 3,304 documents) of the distribution was fitted by Lotka with an estimated α of 1.73. The highly skewed concentration of documents produce that 20 percent (155) of the repositories concentrate circa 90 percent of the documents.

Table 1 presents the summary of the findings about the size of the different types of repositories. From the Average Size column, the first conclusion that can be extracted is that the size of a repository is directly related to its type. The MannÐWhitney U test was applied to corroborate the significance of these differences. The most interesting difference can be found between Learning Object Repositories and Referatories. LORFs are almost an order of magnitude bigger than LORPs. This difference can be explained by the level of ownership required to contribute to these repositories. To contribute to a Referatory, the user only needs to know the address of the learning resource on the Web. Any user can publish any online learning object because its publication does not require permission from the owner of the material. On the other hand, publishing material in a Repository requires, if not being the author of the object, at least to have a copy of it. It can be considered nonethical, or even illegal, to publish a copy of the object without having its ownership or at least the explicit consent of its owner. It will be safe to assume that in the general case, the number of online learning objects that interest a user is larger than the amount of learning objects being authored by herself. Another type of repositories, comparable by its granularity to LORs, is the IRs. The average size, around 10,000, seems to be a midway between the LORF and the LORPs. However, because its power law distribution, IRs could normally be at least two orders of magnitude bigger and 4 order of magnitude smaller than the average. In conclusion, IRs present a larger "range," with some IRs 10 times bigger than the biggest LORF and others 10 times smaller than the smallest LORP. In the large granularity group, OCWs and LMS, the difference is less significant. This similitude can be explained because OCWs are not more than the content of LMS published and made available. The largest and smallest OCWs and LMSs have also similar amount of courses.

If we consider the distribution of the sizes among different types of repositories (Table 1, forth and fifth columns), it is clear that the size of OCWs, LMSs, and IRs is distributed according to a power law with an exponent between 1.5 and 2. This distribution, as mentioned above, produces a wide range of sizes. The variance of the Lotka distribution for these values of $alpha$ is infinite, meaning that it is possible, at least theoretically, that extremely large repositories exist. Also, this distribution presents a heavy tail with few big repositories and a lot of smaller ones. It is surprising that independently of the type of repository, the $alpha$ parameters are similar. On the other hand, LORPs and LORFs present an exponential distribution. However, we argue that in reality, they also follow a Lotka distribution, but the finding of the exponential is an artifact of the sampling method. In the case of OCWs, LMSs, and IRs, the considered repositories were sampled from lists that are not biased to consider only small or large repositories. Any repository, regardless of size, can publish itself in the sampled lists. In the case of LORPs and LORFs, there are no compilation lists available, and the considered repositories are only those that are known, biasing the sample against the expected large amount of small and relatively unknown repositories. An example of this

2. Registry of Open Access Repositories, http://roar.eprints.org/.

sampling artifact can be seen in [27]. There, only the top IRs from ROAR are considered and the size distribution that can be inferred from the presented graphs is distinctively exponential. If all the IRs of ROAR are considered, we found that the real distribution is actually a power law.

The final conclusion that can be extracted from the size analysis is that the distribution of learning objects is very unequal. Most of the resources, independently of the type of repository, are stored in just a few repositories. The concentration is a consequence of the power law distribution. The Pareto or 20/80 rule (also used to describe other heavy-tailed distributions, such as wealth [28]) seems to be a good guide to look at this inequality. The lower value observed in the LORPs and LORFs can also be attributed to the bias toward bigger repositories in the sampling. A more detailed search for LORPs and LORFs will most certainly find small repositories. Adding these small repositories will increase the concentration at the top 20 percent. Despite this inequality distribution, however, no single repository of any type contains more than 40 percent of the available resources. The remaining long tail [29] with the 60 percent of resources is located in other repositories. This can be seen as a strong empirical corroboration of the need to interconnect repositories, either through query federation [30] or metadata harvesting [26].

3 GROWTH ANALYSIS

In order to understand how repositories grow over time, this section analyzes several repositories of different types. The repository growth will be considered in two dimensions: growth in number of objects and growth in the number of contributors. The following sections will present the analysis for each one of these dimensions.

3.1 Content Growth

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 availability of the object publication date was also a determinant factor. The selected repositories are:

- LORPs: ARIADNE,[3] Maricopa Learning Exchange,[4] and Connexions.
- LORFs: INTUTE,[5] MERLOT,[6] and FerlFirst.[7]
- OCWs: MIT OCW and OpenLearn.
- LMSs: SIDWeb.
- IRs-Large: PubMed,[8] Research Papers in Economics,[9] and National Institute of Informatics.[10]
- IRs-University: Queensland,[11] MIT,[12] and Georgia Tech.[13]

3. ARIADNE Foundation, http://www.ariadne-eu.org.
4. Maricopa Learn Exchange, http://www.mcli.dist.maricopa.edu/mlx.
5. INTUTE, http://www.intute.ac.uk.
6. MERLOT, http://www.merlot.org.
7. FerlFirst, http://ferl.becta.org.uk (decommissioned).
8. PubMed repository, http://www.ncbi.nlm.nih.gov/pubmed/.
9. Research Papers in Economics repository, http://www.repec.org.
10. National Institute of Informatics repository, http://www.nii.ac.jp.
11. Repository of U. Queensland, http://espace.library.uq.edu.au/.
12. Repository of MIT, http://dspace.mit.edu/.
13. Repository of Georgia Tech, http://smartech.gatech.edu/dspace.

The collection of data for all the LORs, except INTUTE, consisted in obtaining the date of publication of all their objects. In the case of INTUTE, a sample with all the objects containing the word "Science" (approximately 10 percent of the repository) was obtained. This restriction was set to limit the number of objects to be analyzed due to memory restrictions in the statistical software packages. The data for LORPs and LORFs were collected through Web scraping of the sites during the period between the 5th and the 8th of November 2007. In the case of OCWs and LMSs, the data of publication of all the courses were obtained through direct download. Finally, the selection criteria for the first three IRs (IRs-Large) were size, time of existence, and current activity. These three factors were evaluated from the data provided by ROAR. The second three (IRs-University) were selected from the University repositories of intermediate size with at least three years of existence. The monthly size of these repositories was obtained from data provided by ROAR.

The first variable analyzed was the average growth rate (AGR), measured in objects inserted per day. This value is obtained by dividing the number of objects in the repository by the time difference between the first and last publications. Results for this calculation can be seen in the fourth column (AGR) of Table 2. It is interesting to compare the AGR of different types of repositories. LORPs, for example, grow with a rate of one or two objects per day. OCWs and LMSs grow similarly with an unexpectedly high value of circa one course published per day. From the previous analysis on course size, that rate can be translated, on average, in 20 objects per day. In LORFs and IRs, the variability is significantly higher. For example, big IRs grow more than 10 times faster than University IRs. This difference can be explained by the fact that big IRs are open to a wider base of contributors. On the other hand, the contributor base of University IRs is often restricted to researchers and students of that specific University. The difference between LORFs, however, could not be attributed to size of the contributor community, but to their dedication. INTUTE is a project that pays expert catalogers to find and index learning material on the Web. Merlot and FerlFirst, however, rely on voluntary contributions from external users. A group of paid workers are expected to have a higher production rate than a group of volunteers of a similar size.

The AGR describes linear growth. To test the actual growth function, six models were fitted against the data: linear ($at + b$), biphase linear with breakpoint ($a_1 t$ for $t < Breakpoint$ and $a_2 t + b_2$ for $t \geq Breakpoint$), biphase linear with smooth transition ($ln(a * exp(bx) + c)$), exponential ($b * e^{at}$), logarithmic ($b * ln(at)$), and potential ($b * t^a$). These models were selected based on visual inspection of the size versus time plot (Fig. 2). We use Generalized Linear Model fitting with Least-Squares Estimation. The selection of the model was based on the Akaike information criterion (AIC) [31] that not only takes into account the estimation power of the model, but also its simplicity (less estimated parameters). The result of the fitting indicates that most data sets were best explained by the linear biphase model (both the breakpoint and smooth versions). In PubMed, Connexions, and OCW, the growth is best explained by the potential function, but biphase linear is the second best. A visual inspection of the

TABLE 2
Results of the Growth Analysis of the Repositories

| Repository	Objects	Age (y)	AGR (o/d)	IGR (o/d)	MGR (o/d)	BP (y)
LORP						
Ariadne	4,875	12.0	1.1	2.9	0.66	1.0
Connexions	5,134	7.9	1.8	0.8	2.19	2.8
Maricopa	2,221	4.2	1.4	0.9	2.32	3.0
LORF						
Intute	120,278	12.5	26.7	5.8	36	4.5
Merlot	18,110	10.8	4.6	0.9	5.8	2.8
Ferl First	3,938	6.3	1.7	5.0	1.1	1.0
OCW						
MIT OCW	1,796	4.9	1.0	0.1	2.44	2.9
OpenLearn	499	1.8	0.7	0.1	4.12	1.5
LMS						
SIDWeb	1,445	5.7	0.6	0.2	2.21	4.6
IR						
PubMed	1,124,197	7.3	431	111	591	2.8
RePEc	514,636	4.9	306	65	90	3.3
NII	179,153	5.7	88	42	151	3.6
Queensland	12,069	5.3	6.4	2.4	14	2.7
MIT	27,416	3.8	17.5	32	11	1.6
Georgia Tech	23,163	3.7	20.3	7.4	25	1.4
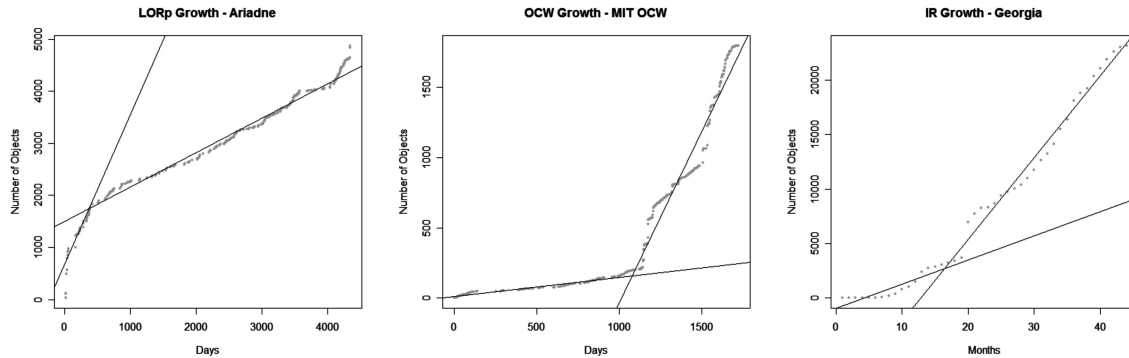
Fig. 2. Empirical and fitted size growth.

plots (Fig. 2) shows indeed that in most data sets, two regimes of linear growth can be easily identified, sometimes with a clear transition point (BP). This result suggests that growth is mainly linear, but the rate is not constant. Two different growth rates are identified in all the repositories. There is an initial growth rate (IGR) that is maintained until a "Breakpoint" (BP) is reached, and then, a mature growth rate (MGR) starts. Table 2 reports the growth rates and breakpoint values for all the studied repositories.

In most cases, the change between IGR and MGR is positive, meaning that the rate increases with maturity. The most logical explanation is that at some point, in time, the repository reaches a critical mass of popularity and the contributor base starts to grow faster, and therefore, the total production rate increases. This hypothesis is tested in the following section when the contributor base growth is studied. However, in two LORs, the production rate decreases from IGR to MGR (Ariadne and FerlFirst). Having inside knowledge of Ariadne history, the inflection point represents the moment when the focus from the Ariadne community shifted from evangelization to attract new members toward interconnection with other repositories through the GLOBE consortium,[14] decreasing the number of

active submissions to the core repository. As such, Ariadne is moving from primarily being a repository to primarily being an integrator of repositories. For FerlFirst, on the other hand, the decline is explained by the abandoning of the project. At the time of writing, this repository has been decommissioned and absorbed by another project, Excellence Gateway.[15] These two can be considered special cases, where the norm is an increase at maturity.

Another interesting finding is that BP is, in most cases, located between two and three years after the first object has been inserted into the repository. In the case of LORs, this can be seen as the time needed by the repository to reach a critical mass of objects that could attract more users or funding, and therefore, more objects. In the case of LMSs, OCWs, and University IRs, this can be the time taken to "cross the chasm" [32] between early adopters and mainstream use inside the institution.

It is important to note that the linear trend is observed at large time scales. The short-term growth, specially for OCWs, LMS, and IR, is characterized by irregular "jumps." These jumps can be explained by external events such as the start of the academic semesters or the deadline for annual reviews. If we smooth these jumps over a long period of time; however, the linear growth is apparent.

TABLE 3
Result of the Analysis of the Contributor Growth in the Repositories

Repository	Contrib.	AGR (c/d)	Function	Growth Rate	BP (y)
LORP					
Ariadne	166	0,04	Bi-Phase L.	IGR=0.02, MGR=0.06	3.5
Connexions	581	0,22	Exponential	ER=1.2 x 10-3	3.0
Maricopa	529	0,20	Bi-Phase L.	IGR=0.06, MGR=0,28	2.4
LORF					
Merlot	1,446	0.42	Bi-Phase L.	IGR=0.12, MGR=0.54	1.1
OCW					
MIT OCW	1,072	1.34	Exponential	ER= 3.7 x 10-3	2.3
LMS					
SIDWeb	584	0.21	Exponential	ER=1.8 x 10-3	3.3
IR					
Queensland	9,503	3.9	Bi-phase L.	IGR=1.6, MGR=6.9	3.8
MIT	21,028	9.7	Bi-phase L.	IGR=0.7, MGR=17	2.2
Georgia Tech	10,704	7.9	Bi-phase L.	IGR=4.4, MGR=9.7	1.7

The main conclusion of this analysis is that linear growth in the number of objects is a sign of the lack of penetration of Learning Object technologies in educational settings. It would be expected that the amount of learning material created follows an exponential growth [33]. It seems that much of this material is not published in any type of repository.

3.2 Contributor Base Growth

Another way to consider repository growth is to measure its contributor base at different points in time. For this analysis, we try to use the same set of data as in the previous experiment, with some exceptions. Intute, FerlFirst (LORFs), and OpenLearn (OCW) were excluded because they do not provide contributor data for their objects. To obtain the data for the IRs, the complete metadata set was harvested from the repositories. For this reason, the three biggest IRs (PubMed, RePEc, and National Institute of Informatics) were excluded. It was not feasible to process the large amount of information using a single computer in a reasonable amount of time. Nonetheless, the remaining repositories are representative of their respective type.

The list of contributors was obtained by taken the name of the authors of all the objects present in the repository. A list of all the unique contributors was generated. In the case that one object have several authors, only the first author is considered its contributor. For example, if the learning object metadata mention three authors, the act of publication is only attributed to the first author. Counting the first author instead of assigned fractional counts to all the authors is a common practice in Scientometrics [34].

As can be seen in the second column of Table 3, the size of the current contributor base of the studied repositories (with the exception of IRs) is within one order of magnitude. The smaller contributor base is ARIADNE (166) and the largest is Merlot (1,446). It is important to note that the difference in numbers of objects of LORP and LORF cannot be explained just by the size of the contributor base. As it was hypothesized in the previous section, the difference should originate in a different rate of contribution. In the case of IRs, the user base is considerably bigger than in other types of repositories. The reason for this difference can be found in the fact that contributing to

Institutional Repository is, in most cases, mandatory for postgraduate students [35]. Publishing their thesis in the Institutional Repository is commonly a requirement. Contrarily, contribution to the other type of repositories is optional, and in the case of OCWs and LMSs, it is normally reserved for professors only.

The analysis of the AGR, in the third column of Table 3, presents an even more coherent picture. The number of authors in LORs seems to incorporate a new contributor each two to five days. In MIT OCW, with its accelerated publishing program, a new professor is added every day on average. SIDWeb has a rate similar to LORs with a new professor using the system every five days, also on average. As expected from the previous analysis, IRs have a much higher AGR, counting between four and 10 new contributors each day. Based on these numbers, it would be an interesting experiment to open LMSs and OCWs also to under- and postgraduate student contributions.

The actual growth function was determined with the same candidate functions and fitting procedure used in the previous section. The results can be seen in Table 3, fourth and fifth columns. While the majority of repositories present a bilinear growth, similar to the growth in the number of objects, it was surprising to find that in the case of MIT OCW, SIDWeb, and Connexions, the contributor base grows exponentially. This effect can be better visualized in Fig. 3. Section 5 will present a model that could explain how an exponentially growing user base could generate a linear object growth.

The change from IGR to MGR is positive in all the studied repositories. In most cases, the MGR is between two and four times larger than the IGR. A notable exception is MIT IR. In that particular case, the rate increases more than 10 times. In the case of exponential growth (Connexions, MIT OCW, and SIDWeb), the rate at which new users enter the system is always increasing. That growth is captured in the exponential rate parameter. This parameter is similar for the three repositories, with MIT OCW also presenting the most rapid growth. However, exponential growth cannot continue forever, especially in the case of MIT OCW and SIDWeb. Once most of the professors have created a course in those systems, the contributor growth rate will follow the
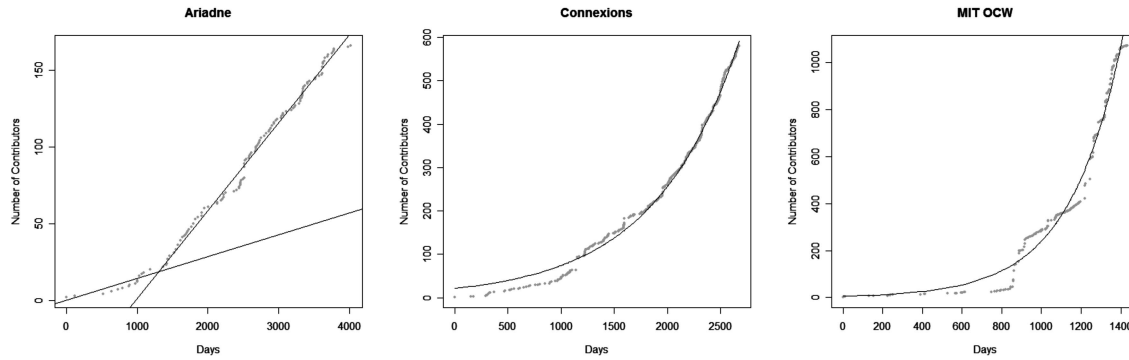
Fig. 3. Empirical and fitted contributor base growth.

TABLE 4
Analysis of Distribution of Contribution

Repository	AC (o)	Distribution	Parameters	C-20 (%)
LORP				
Ariadne	29.4	Lotka exp. cut-off	$\alpha = 1.57, \lambda = 0.011$	75%
Connexions	8.84	Lotka exp. cut-off	$\alpha = 1.35, \lambda = 0.0094$	78%
Maricopa	4.20	Lotka exp. cut-off	$\alpha = 2.12, \lambda = 0.0067$	64%
LORF				
Merlot	12.5	Lotka exp. cut-off	$\alpha = 1.88, \lambda = 0.0006$	82%
OCW				
MIT OCW	39.7	Weibull	$k = 1.07, \lambda = 40.5$	50%
LMS				
SIDWeb	40.0	Weibull	$k = 0.52, \lambda = 17.14$	72%
IR				
Queensland	1.3	Lotka	$\alpha = 3.01 , xmin = 2$	22%
MIT	1.3	Lotka	$\alpha = 2.55. xmin = 3$	36%
Georgia Tech	2.1	Lotka	$\alpha = 2.25, xmin = 5$	68%

incorporation of new faculty members, which is commonly linear over large periods of time.

An interesting comparative analysis can be made between the breakpoints of the object and contributor base growth. This analysis can shed light on the "chicken or egg" dilemma regarding repositories. This dilemma can be summarized as: Does an increase in the number of objects attract more users? Or does the increase in the number of contributors generate more objects? The breakpoint in the case of Biphase linear growth is the point of transition between the two linear phases. In the case of exponential growth, we selected the point where the function grows faster than linear. Comparing the BP values in Tables 2 and 3, the evidence is inconclusive. Some repositories, such as Maricopa, MER-LOT, and MIT OCW, first have an increase in the rate at which new contributors arrive and eight months later, on average, an increase in rate of growth is perceived. However, in the IRs, SIDWeb, and Connexions, the contrary is true. First, the rate object growth increases, and four months to one year afterward, the rate of new contributors follows. It seems that to gain insight on how the "chicken or egg" dilemma is solved in the repositories, a deeper analysis is needed.

A final conclusion for this analysis brings hope for the establishment of Learning Object Technologies. Spotting exponential growth in the number of contributors in three repositories is maybe signaling that a transition phase between linear and exponential growth is happening. However, this change will depend on the ability of the repositories to retain the productive users. This will be ultimately defined by the engagement and fidelity that the repository could produce in their contributors.

4 CONTRIBUTION ANALYSIS

In this section, we analyze in more detail how contributors publish objects in the repositories. The first study will analyze how many objects are published by each contributor. The second section analyzes how frequently objects contributors publish objects. The third and last section examines the amount of time that each contributor keeps contributing objects. These analyzes will help us to gain an insight on the inner workings of the learning object publication process.

4.1 Contribution Distribution

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.

The first step in this analysis is to obtain the average number of publications per contributor (AC). This value was obtained dividing the total number of objects in the repository (Table 2) by the number of contributors (Table 3). Table 4 presents this value in the second column. It is interesting to note that the average output of contributors to different kind of repositories differ substantially. The contributors to the ARIADNE repository, while few, have
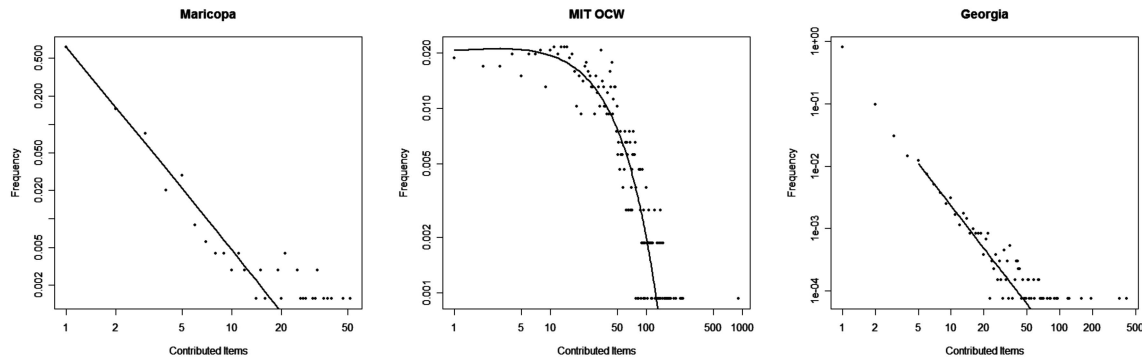
Fig. 4. Empirical and fitted distribution of the number of publications between contributors.

produced, during the lifetime of the repository, more objects per capita than any of the other LORs. The results for Merlot also confirm that the bigger size of LORF is not due to a bigger, but to a more productive contributor base. The results for MIT OCW and SIDWeb show that they have a similar productivity per contributor, hinting that the publishing mechanics in LMSs and OCWs are very similar. For IRs, the only way to explain the low productivity, in what it is a scientific publication outlet, is to assume that the majority of the content is made by student thesis.

The next step in the analysis is to obtain an approximation for the distribution of the number of publication for each author. Given that the data are highly skewed to the left, the best way to present it is the size-frequency graph in logarithmic scales (Fig. 4). This figure represents how probable (y-axis) is to find a contributor that has published a certain amount of objects (x-axis). Five statistical distributions were fitted against the data: Lotka, Lotka with exponential cutoff, Exponential, Log-Normal, and Weibull. The parameter estimation was made with the MLE method and the Vuong test was used to find the best fitting of the competing distributions. The best-fitting distribution and their estimated parameters can be seen in the third and fourth columns of Table 4.

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 LORs follow a Lotka distribution with exponential cutoff. The meaning of this cutoff is that it becomes increasingly harder to publish a large amount of objects. The effect can be seen in Fig. 4 as a slight concavity at the tail of the distribution. The parameters for the different LORs are similar. Even the cut off rate seems to agree. For LORPs, the cutoff starts sooner by an order of magnitude than for LORF. The finding of this distributions means that most LOR contributors only publish one object. Even high producing individual starts loosing interest after publishing many objects. Maybe one of the reasons behind this distribution is the lack of some type of incentive mechanism [7].

OCW MIT and SIDWeb present a Weibull distribution. The finding of this 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 one object per session).

The tails of the IRs are fitted by the pure Lotka distribution. The head of the distribution and users that have published one or two objects have a disproportionately high value that cannot be fit by any of the tried distributions. The tails, however, have an α of around 2.50 that is consistent with previous studies [36] [37] of the distribution of scientific publications among authors. This result suggests that the publication of documents in IRs has a different mechanism than the publication of learning objects in LORs, and maybe what we are measuring in the IRs tail is a byproduct of the scientific publication process.

Finally, the percentage of objects created by 20 percent of the users is calculated. The results are presented in the last column of Table 4 (C-20). From these results, it can be concluded that the LORs are affected by the Pareto inequality (20/80 rule). The concentration for OCW MIT is less unequal, with just 50 percent of the objects being published by the top 20 percent contributors. The explanation for this result is that there is a considerable proportion of users that produce between 10 and 50 objects. This group of contributors is productive enough to balance the production of the tail. In the case of IRs, an interesting effect, produced by the thesis publication, can be observed. The most productive section of the contributors is located at the head of the distribution. This effect is more visible in the Queensland repository. There, 20 percent of the most productive contributors also publish around the 20 percent of the material. The only way to reach this percentage is that almost all the contributors publish the same amount of documents, more concretely for this repository, one document.

As can be concluded from the previous analysis, all contributors are not equal. In any population, regardless if it is Weibull or Lotka, there will always be several "classes" of users, similar to the segmentation use to classify socio-economic strata (as mentioned before, income is also heavy-tailed distributed). We can divide the contributing population in a large "lower class" of contributors that only publish few objects. A smaller "middle class" that publishes intermediate amount of objects and a very small "higher class" that publishes a large amount of objects. These classes arise naturally and have to be dealt with. While the publishing capacity can be increased with better tools and intuitive environments, the inherent inequality will, most probably, persist.

TABLE 5
Result of the Analysis of Lifetime

Repository	ALT (d)	LT Distribution	LT Parameters
LORP			
Ariadne	514	Exponential	$\lambda = 0.0010$
Connexions	261	Exponential	$\lambda = 0.0012$
Maricopa	304	Exponential	$\lambda = 0.0012$
LORF			
Merlot	328	Exponential	$\lambda = 0.0015$
OCW			
MIT OCW	67	Weibull	$k = 1.72$, $\lambda = 325$
LMS			
SIDWeb	364	Weibull	$k = 1.21$, $\lambda = 588$
IR			
Queensland	319	Log-Normal	$\mu log = 6.01$, $\sigma log = 0.89$
MIT	120	Log-Normal	$\mu log = 3.84$, $\sigma log = 2.45$
Georgia Tech	9.3	Log-Normal	$\mu log = 3.61$, $\sigma log = 2.16$

TABLE 6
Result of the Analysis of Publishing Rate

Repository	APR (o/d)	Rate Distribution	Rate Parameters
LORP			
Ariadne	0.082	Log-Normal	$\mu log = -3.25$, $\sigma log = 1.27$
Connexions	0.046	Log-Normal	$\mu log = -4.11$, $\sigma log = 1.36$
Maricopa	0.010	Log-Normal	$\mu log = -5.18$, $\sigma log = 0.95$
LORF			
Merlot	0.37	Log-Normal	$\mu log = -2.47$, $\sigma log = 1.11$
OCW			
MIT OCW	0.32	Log-Normal	$\mu log = -1.68$, $\sigma log = 1.07$
LMS			
SIDWeb	0.12	Log-Normal	$\mu log = -2.57$, $\sigma log = 0.96$
IR			
Queensland	0.17	Log-Normal	$\mu log = -4.05$, $\sigma log = 2.07$
MIT	0.24	Log-Normal	$\mu log = -7.08$, $\sigma log = 2.14$
Georgia Tech	1.55	Log-Normal	$\mu log = -1.84$, $\sigma log = 2.53$

ALT is measured in days.

4.2 Lifetime and Publishing Rate

In this analysis, we consider the publication history of each individual contributor. Two variables will be measured: First, the lifetime of the contributor. This is the time from its first to last publication. Second, the publishing rate. This is the average number of objects published during the lifetime of the contributor. In real-world terms, the lifetime can be considered as the period during which the contributor is engaged with the repository. The publishing rate, on the other hand, can be considered as proxy measurement of the talent or capacity that the contributor has to publish learning objects. To compute these values, we extract the contributor information from the repositories used in the previous analysis.

In the calculation of the lifetime, we always know its beginning, but we are never sure about its end. A contributor could have published its first object two years ago and its last object one year ago. The measured lifetime will be one year. However, if the contributor published one more object just the day after the data were captured, its actual lifetime will be two years. To cope with this limitation, the lifetime of a user is only considered finished if the time from the last object insertion is at least as long as the longest period without activity between two consecutive publications. If a lifetime is not ended, it will be assigned the time interval from the first object insertion until the date of data collection.

The measurement of the rate of production also presents some difficulties. The rate of contribution could not be measured if all the objects have been published on the same day. Also, the publication of few objects in a short lifetime will produce inflated rate values. To alleviate this problem, only users whose lifetime is larger than 60 days, and that have published at least two objects are considered for the calculation. To prevent the bias in the distribution produced by only considering highly productive contributors, the contributors that have a lifetime shorter than 60 days are assigned the smallest production rate.

As expected, each contributor present a different lifetime. To obtain a clearer picture of how these values compare between different repositories, we calculate the Average Lifetime (ALT). Table 5 presents the results measured in days. The first conclusion that can be extracted from the lifetime values is that, on average, they are much smaller than the lifetime of the repositories (Table 2). This means that most contributors are "retired" after a period of one year. This conclusion holds even if the contributors "born" during the last year are removed from the calculation. However, the actual values of ALT are not related to the type of repository and do not provide much information about the distribution of the lifetime among contributors.

Based on the skewed nature of the distribution of lifetime, we fit the five heavy-tailed distributions used previously. The results can be seen in Table 6. The

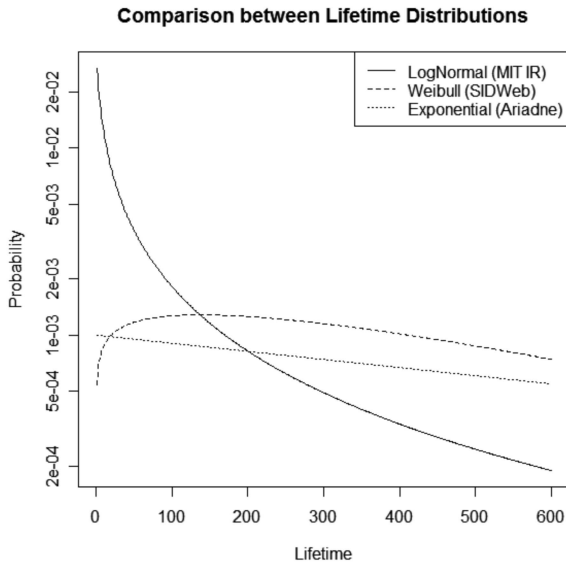**Comparison between Lifetime Distributions**

Fig. 5. Comparison between lifetime distribution between repository types.

distribution is clearly related to the type of repository. A comparison between the different lifetime distributions across several repository types can be seen in Fig. 5.

LORs contributors have lifetimes that are distributed exponentially among the population. The λ parameter of the exponential is also similar across LORs. This similarity suggests that LORs contributors share the same type of engagement with the repository. The probability of cease publishing is proportional to the time that the contributor has been active. The result is that there are considerable amount of users with short lifetimes (less than three months). We can classify this behavior as engagement by novelty. As the novelty worn off, the user ceases contributing.

In the case of OCWs and LMS, the lifetime follows a Weibull distribution. Again, the shape and scale parameters are similar. A Weibull distribution with these parameters hints that the amount of contributors with very short lifetimes (less than one month) does not dominate the population. It is more common to find contributors that keep publishing after three months to one year. However, Weibull decreases rapidly after its peak, meaning that it is infrequent to find contributors with several years of publication. We describe this behavior as engagement by need. The average contributor keeps publishing until a goal is reached (for example, a course is completed and/or improved). After the goal has been reached the probability of stopping increases rapidly.

The IRs present a very different publishing behavior, denoted by the Log-Normal distribution. The low σlog parameter, found in all the IRs lifetimes, means that the majority of the contributors have a very short lifetime (few weeks), with a neglectable amount having lifetimes measured in months or years. This result is consistent with the finding that most contributors in the studied IRs published just one object, probably their thesis. After this publication, and maybe some immediate corrections or additions, the contributors cease publishing. We describe this behavior as low engagement. The repository does not require or

promote continuous submissions from the majority of users. As a response, most user lifetimes are basically instantaneous compared with the lifetime of the repository.

After analyzing the lifetime, we calculate the average publication rate (APR) for each repository. The results are presented in Table 6. The only clear conclusion that can be extracted from the APR is that LORP contributors publish less frequently than other types of repositories. Specially, if compared with the similar LORF, MERLOT, the publication rate seems to be one order of magnitude lower. As mentioned before, the main difference between the size of LORPs and LORFs seems to indicate a difference in productivity of the contributor base. The difference of APR between the other type of repositories is not clear.

To gain better insight on how the publication rate is distributed across the contributing population, the five previously used statistical distributions are fitted to the data. The results of the fitting are presented in Table 6. Surprisingly, all the repositories show the same distribution, Log-Normal. The main difference seems to be that the σlog parameter is around one for LORPs, LORFs, OCWs, and LMSs, and around 2 for IRs. A higher σlog creates a larger skewness to the left, meaning that a larger proportion of contributors is low-productive. The finding of the same distribution for all the repositories is very significant, because it means that there is no difference between the distribution of talent or capacity among the different contributor communities.

This analysis shows that the main differentiator between different types of repositories is the type of engagement that the contributors have. According to the findings, the most successful models of repository seems to be OCWs and LMSs, where most of the contributors keep publishing for longer periods of time. This result suggests again that an incentive-based publishing is the most effective form to increase the total number of learning objects available.

5 IMPLICATION OF THE FINDINGS

The results of the quantitative analysis can be used to answer the questions raised in Section 1. This section presents those answers and the implications that they have in our understanding of the learning object publication process and the technological design of repositories.

What is the typical size of a repository? Is it related to its type? In general, individual learning object repositories seems to vary from hundreds to millions of objects. Their average size depends on 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, these 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 thousands of courses.

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. Sampling biases aside, 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 [30] 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 [26] 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.

How repositories grow over time? Linearly. 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 exponentially, it is not able to retain them long enough to feel the effect. The value equation, how the contributor benefits from contributing to the repository, is still an unsolved issue in most repositories. Several researches have suggested incentive mechanism [11] [40] comparable to scientific publication, in order to provide the professor with some type of reward for their contribution.

Another interesting result in the growth analysis was to find that all repositories went through an initialization with usually a very low growth rate. The length of this stage varied from one to three years (shortening for more recent repositories). After this period, a more rapid expansion begins caused by (or that cause) an increase in the number of contributors joining the repository. Having knowledge of these phases could help repository administrators to not discard slow growing repositories too soon.

What is the typical number of contributors a repository has? Is it related to its type? We can estimate, from the analysis in Section 3, that medium LORs have a base of 500 to 1,500 contributors. This number is also similar for OCWs and LMSs contributor bases. On the other hand, IRs, being targeted also to students, have contributor bases one order of magnitude bigger. The size of the contributor base, however, is not always related to the size of the repository. Merlot contributors, being outnumbered 1-10, produce a comparable amount of objects as MIT IR contributors. Moreover, the title to the most productive contributors in the study goes to the OCWs and LMSs professors (Table 4) with around 40 objects on average. This results also support the idea that LMSs are the most effective type of repository, given that they provide a clear value into the publishing step (students not asking for copies of the material, for example).

Given the relatively small size of the communities that build repositories, it would be an interesting experiment to measure the impact that the introduction of social networks could have in the sharing of material. For example, users would be interested in knowing when a colleague in his same field has published new learning objects [41]. This social networks can be created explicitly (a lá Facebook) or implicitly (relationship mining) [42]. The deployment of these types of networks could also help to solve the lack of engagement problem.

How the number of contributors grows over time? Most of them linearly, but surprisingly three of them Connexions, MIT OCW, and SIDWeb, exponentially. This unexpected result, specially in SIDWeb, a run-of-the-mill LMS, is very encouraging for the future of the Learning Object Economy, because it can give rise, with the right environment, to exponential growth of content available. However, we also found that at this stage, the growth in these repositories continues linear. However, this observation can be due to the recent kickoff of exponential contributor base growth in these repositories. A follow-up study in a year period would help us to have a better perspective. Again, the finding of exponential growth in course-based repositories confirms the idea that we should strive to connect LMS as the main source of learning material.

How many learning objects a contributor publishes on average? As mentioned before, the average productivity of users depends on the type of repository. For LORP, it can be around 10 objects per contributors. IRs present the lowest production per contributor with 1 or 2 on average. However, heavy tail distributions, Lotka and Weibull, make this answer a little more complicated. The problem with the average values given previously in the current situation is that in heavy tailed distributions, "there is not such thing as an average user." As mentioned in Section 4, 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 high class. Most of the material is created by a few hyperproductive contributors. the 10 percent 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 classes are comparatively small. Finally, University IRs, with Lotka with high $alpha$, are dominated by the lower class as more than 98 percent of the population produce just one object. From our analysis on publishing rate and lifetime, we can conclude that these different distributions are caused not 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 keeps the contributor active at the beginning, but the chances of ceasing publication increase as more time is spent in the repository. For OCWs and 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.

In conclusion, it is very important for a repository administrator to know the composition and characteristics of her contributor base. Having a clear view of what and

who need to be incentivized is the first step before building any type of incentive plan [43].

It is also interesting to note that these distributions are not exclusive for the publication of learning objects, but are shared by various types of user-generated content (UGC) [44] . Moreover, there is a long research history of how the Lotka law fits the process of scientific publishing [37]. The quantitative study of Learning Objects (or Learnometrics as we call it) can borrow substantial amount of research results from other Informetrics fields. Moreover, Informetrics could also benefit from having new sources of data in an specific domain to test the generality of their conclusions.

Despite the previous answers, this analysis raises more questions than it solves. We invite the reader to check the Further Research section at the end of this dissertation to share what we consider to be the most interesting new paths opened by this work.

6 CONCLUSIONS

This paper is the first quantitative analysis performed to the publication of learning objects. We have raised and answered several basic questions important for the understanding of the publication process and the design and operation of learning object repositories of several types.

Maybe the most relevant conclusion from the quantitative analysis is that the publication process is dominated by heavy-tailed distributions and the usual Gaussian-based statistics are not enough to gain insight on the nature of the compiled data. These distributions also provide the repositories with several characteristics not found in more normal sets. For example, difference in size or productive can span through several order of magnitude. Depending on the parameters of the distributions, it will not be unexpected that most of the content of a repository is produced but few individuals or that 99 percent contributor base only publish one object. The black swan effects [45] can be seen, measured and modeled in the composition of all repositories.

Finally, measuring the publication process enables us to take better decisions about the architecture and infrastructure needed to support the Learning Object Economy. Moreover, measuring is our only way to test the unproven assumptions over which some of the current Learning Object technology rests.

To complement this study about the supply of learning objects, the next paper will analyze the other side of the economy: the demand. Having a clear view of how these two process work could help Market-Makers and Policy-Makers to understand how different technologies and policies affect the Learning Object Economy.

7 FURTHER RESEARCH

The quantitative analysis of the production of learning objects has received little research attention, although it is the base of Learning Object Economy. This paper answered some basic questions, but much more are left open. Moreover, embedded in the provided answers, there are the seeds of new questions.

- Effect of openness. An interesting question is whether repositories with open publication, such as Connexions or Merlot, are more efficient or productive that closed projects, such as Intute or MIT OCW, in the long run.
- Publication patterns. While the easy metrics of lifetime and rate of contribution give an idea of the publication process, a deeper analysis of the patterns in which publications take place would add more information to understand this process.
- How to integrate LMS. One of the interesting findings of this paper is that LMSs seem to be the best environment for learning object publication. However, traditionally, these are isolated silos of information. How to intercommunicate between them and share their contents, not only at the technical level, but also, and more importantly, at the social, legal, and administrative level should be one of the main challenges of our field.

Finally, we make a call for more quantitative analysis for Learning Objects. There are other important aspects in the life cycle of the learning objects that we know little about: creation, reuse, versioning, etc. The fact that the data for this kind of studies are hard to get does not make the questions that those measurements could solve less important or urgent. Measuring will let us know where we are and whether we, indeed, are moving forward.

REFERENCES

[1] B. Collis and A. Strijker, "Technology and Human Issues in Reusing Learning Objects," *J. Interactive Media in Education,* vol. 4, pp. 1-32, 2004.

[2] E. Duval et al., "The Ariadne Knowledge Pool System," *Comm. ACM,* vol. 44, no. 5, pp. 72-78, 2001.

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

[4] T. Malloy, G. Jensen, A. Regan, and M. Reddick, "Open Courseware and Shared Knowledge in Higher Education," *Behavior Research Methods, Instruments, and Computers,* vol. 34, no. 2, pp. 200-203, 2002.

[5] T. Malloy and G. Hanley, "MERLOT: A Faculty-Focused Web Site of Educational Resources," *Behavior Research Methods, Instruments, and Computers,* vol. 33, no. 2, pp. 274-276, 2001.

[6] A. Agogino, "Visions for a Digital Library for Science, Mathematics, Engineering and Technology Education (SMETE)," *Proc. Fourth ACM Conf. Digital Libraries,* pp. 205-206, 1999.

[7] L. Campbell, "Engaging with the Learning Object Economy," *Reusing Online Resources: A Sustainable Approach to E-Learning,* pp. 35-45, Kogan Page, Ltd., 2003.

[8] G. Richards, R. McGreal, M. Hatala, and N. Friesen, "The Evolution of Learning Object Repository Technologies: Portals for On-Line Objects for Learning," *J. Distance Education,* vol. 17, no. 3, pp. 67-79, 2002.

[9] S. Ternier, "Standards Based Interoperability for Searching in and Publishing to Learning Object Repositories," PhD dissertation, Katholieke Univ. Leuven, 2008.

[10] S. Guth and E. Köppen, "Electronic Rights Enforcement for Learning Media," *Proc. IEEE Int'l Conf. Advances Learning Technologies (ICALT '02),* V. Petrushin, P. Kommers, Kinshuk, and I. Galeev, eds., pp. 496-501, Sept. 2002.

[11] S. Downes, "Models for Sustainable Open Educational Resources," *Interdisciplinary J. Knowledge and Learning Objects,* vol. 3, pp. 29-44, 2007.

[12] R. McGreal, "A Typology of Learning Object Repositories," http://hdl.handle.net/2149/1078, 2007.

[13] S. Retalis, "Usable and Interoperable E-Learning Resources and Repositories," *Interactive Multimedia in Education and Training,* Idea Group, 2004.

[14] A. Tzikopoulos, N. Manouselis, and R. Vuorikari, "An Overview of Learning Object Repositories," *Learning Objects for Instruction: Design and Evaluation,* pp. 29-55, Idea Group, 2007.

[15] F. Neven and E. Duval, "Reusable Learning Objects: A Survey of Lom-Based Repositories," *Proc. 10th ACM Int'l Conf. Multimedia (MULTIMEDIA '02),* M. Muhlhauser, K. Ross, and N. Dimitrova, eds., pp. 291-294, 2002.

[16] M. Sicilia, E. Garcia, C. Pages, and J. Martinez, "Complete Metadata Records in Learning Object Repositories: Some Evidence and Requirements," *Int'l J. Learning Technology,* vol. 1, no. 4, pp. 411-424, 2005.

[17] EdReNe, "Current State of Educational Repositories," technical report, eContentplus, 2009.

[18] L. Egghe and R. Rousseau, "Systems without Low-Productive Sources," *Information Processing and Management,* vol. 42, no. 6, pp. 1428-1441, 2006.

[19] M. Goldstein, S. Morris, and G. Yen, "Problems with Fitting to the Power-Law Distribution," *The European Physical J. B-Condensed Matter,* vol. 41, no. 2, pp. 255-258, 2004.

[20] Q. Vuong, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica,* vol. 57, no. 2, pp. 307-333, 1989.

[21] A. Clauset, C. Shalizi, and M. Newman, "Power-Law Distributions in Empirical Data," p. 26, Arxiv preprint arXiv:0706.1062, 2007.

[22] R. Rousseau, "Lotkas Law and Its Leimkuhler Representation," *Library Science with a Slant to Documentation Studies,* vol. 25, pp. 150-178, 1988.

[23] Q.L. Burrell, "The Gini Index and the Leimkuhler Curve for Bibliometric Processes," *Information Processing and Management,* vol. 28, no. 1, pp. 19-33, 1992.

[24] F. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *J. Am. Statistical Assoc.,* vol. 46, no. 253, pp. 68-78, 1951.

[25] J. Cole and H. Foster, *Using Moodle: Teaching with the Popular Open Source Course Management System.* O'Reilly Media, Inc., 2007.

[26] H. Van de Sompel, M. Nelson, C. Lagoze, and S. Warner, "Resource Harvesting within the OAI-PMH Framework," *D-Lib Magazine,* vol. 10, no. 12, pp. 1082-9873, 2004.

[27] L. Carr and T. Brody, "Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles," *D-Lib Magazine,* vol. 13, nos. 7/8, pp. 1082-9873, 2007.

[28] H. Wold and P. Whittle, "A Model Explaining the Pareto Distribution of Wealth," *Econometrica,* vol. 25, no. 4, pp. 591-595, 1957.

[29] C. Anderson, *The Long Tail.* Hyperion, 2006.

[30] 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," *Proc. First Workshop Interoperability of Web-Based Educational Systems,* D. Olmedilla, N. Saito, and B. Simon, eds., pp. 11-18, 2005.

[31] H. Akaike, "An Information Criterion (AIC)," *Math. Sciences,* vol. 14, no. 153, pp. 5-9, 1976.

[32] G. Moore, *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers.* Collins, 1991.

[33] P. Lyman and H.R. Varian, "How Much Information?" *J. Electronic Publishing,* vol. 6, no. 2, p. 8, 2000.

[34] D. Lindsey, "Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship," *Social Studies of Science,* vol. 10, no. 2, pp. 145-162, 1980.

[35] P. Davis and M. Connolly, "Institutional Repositories," *D-Lib Magazine,* vol. 13, nos. 3/4, pp. 1082-1101, 2007.

[36] R. Coile, "Lotka's Frequency Distribution of Scientific Productivity," *J. Am. Soc. for Information Science,* vol. 28, no. 6, pp. 366-370, 1977.

[37] M. Pao, "An Empirical Examination of Lotka's Law," *J. Am. Soc. for Information Science,* vol. 37, no. 1, pp. 26-33, 1986.

[38] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan, and M. Smith, "The DSpace Institutional Digital Repository System: Current Functionality," *Proc. Joint Conf. Digital Libraries,* pp. 87-97, 2003.

[39] C. Lagoze, S. Payette, E. Shin, and C. Wilper, "Fedora: An Architecture for Complex Objects and Their Relationships," *Int'l J. Digital Libraries,* vol. 6, no. 2, pp. 124-138, 2006.

[40] H. Hummel, D. Burgos, C. Tattersall, F. Brouns, H. Kurvers, and R. Koper, "Encouraging Contributions in Learning Networks Using Incentive Mechanisms," *J. Computer Assisted Learning,* vol. 21, no. 5, pp. 355-365, 2005.

[41] E. Duval, "LearnRank the Real Quality Measure for Learning Materials," *Policy and Innovation in Education—Quality Criteria,* pp. 457-463, European Schoolnet, 2005.

[42] Y. Matsuo, J. Mori, M. Hamasaki, K. Ishida, T. Nishimura, H. Takeda, K. Hasida, and M. Ishizuka, "Polyphonet: An Advanced Social Network Extraction System from the Web," *Proc. 15th Int'l Conf. World Wide Web (WWW '06),* pp. 397-406, 2006.

[43] B. Berendt, E. Brenstein, Y. Li, and B. Wendland, "Marketing for Participation: How Can Electronic Dissertation Services Win Authors?" *Proc. Next Steps—Electronic Theses and Dissertations Worldwide (ETD '03),* pp. 156-161, 2003.

[44] X. Ochoa and E. Duval, "Quantitative Analysis of User-Generated Content on the Web," *Proc. First Int'l Workshop Understanding Web Evolution (WebEvolve '08),* D. De Roure and W. Hall, eds., pp. 19-26, 2008.

[45] N. Taleb, *The Black Swan: The Impact of the Highly Improbable.* Random House, 2007.

Xavier Ochoa is a professor in the Faculty of Electrical and Computer Engineering at Escuela Superior Politécnica del Litoral (ESPOL). He coordinates the research group on Technology Enhanced Learning at the Information Technology Center (CTI), ESPOL. He is also involved in the coordination of the Latin American Community on Learning Objects (LACLO) and several regional projects. His main research interests revolve around measuring the Learning Object economy and its impact in learning.

Erik Duval is a professor in the research unit on hypermedia and databases in the Computer Science Department of the Katholieke Universiteit Leuven. He teaches courses on human-computer interaction, multimedia, problem solving, and design. His current research interests are metadata, learning object metadata in particular, a global learning infrastructure based on open standards, and human-computer interaction in the learning or digital repository context. He serves as the copresident of the ARIADNE Foundation, the chair of the IEEE LTSC Working Group on Learning Object Metadata, and a member of the Scientific and Technical Council of the SURF Foundation. He is a fellow of the AACE, and a member of the ACM, the IEEE, and the IEEE Computer Society.