1-1-2011

Cost Factors in Digital Projects: A Model Useful in Other Applications

Lisa L. Crane
Claremont Colleges Library

THE FRUGAL LIBRARIAN

Thriving in Tough Economic Times

Edited by Carol Smallwood

American Library Association
Chicago 2011

Printed in the United States of America
15 14 13 12 11 5 4 3 2 1

ISBN: 978-0-8389-1075-7

Library of Congress Cataloging-in-Publication Data
The frugal librarian : thriving in tough economic times / edited by Carol Smallwood.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-8389-1075-7 (alk. paper)
1. Library finance--United States. 2. Libraries--United States--Cost control. 3. Libraries--Economic aspects--United States. 4. Library science--Economic aspects--United States. 5. Libraries and community--United States. 6. Libraries--United States--Case studies. I. Smallwood, Carol, 1939-
Z683.2.U6F73 2011
025.1'10973--dc22

2010034317

Book design in Charis SIL and Avenir by Casey Bayer.

♾ This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

ALA Editions also publishes its books in a variety of electronic formats. For more information, visit the ALA Store at www.alastore.ala.org and select eEditions.

18

COST FACTORS IN DIGITAL PROJECTS

A Model Useful in Other Applications

Lisa L. Crane

THE CLAREMONT UNIVERSITY Consortium (CUC) is the central coordinating and support organization for seven independent colleges, known as the Claremont Colleges, located in Southern California. A centralized library is one of a myriad of services provided by CUC to the colleges. The library contains a digital production unit consisting of four full-time staff and several part-time student workers. The output of the unit is the Claremont Colleges Digital Library (CCDL). Established in April 2006, the CCDL provides the infrastructure to disseminate unique resources held by the Claremont Colleges and the Claremont Colleges Library.

In March 2009, the digital production unit was asked to provide the cost to put an item into the CCDL. Fortunately, the unit had been gathering data since the fiscal year began on July 1, 2008. Now there was an impetus to crunch the numbers. It took the digital production librarian about two full weeks sequestered behind closed doors, doing nothing but number crunching.

What follows in this case study is not a complete instruction on cost accounting. Rather, I introduce concepts and share the tools and methods used. The numbers in this case study are actual numbers and cover the period from July 1, 2008, through the end of February 2009.

THE VARIABLES

There are two major cost variables when it comes to digital projects: material type and funding/wages. As with most digital libraries, or digital projects, the CCDL contains a variety of source material that must be digitized, uploaded, and described through metadata, including photographs, glass plate negatives, lantern slides, 35 mm slides, videos, oversized materials, scrapbooks, monographs, and documents.

Costs also depend on the funding provided and the various wages paid to students. At the time of our cost analysis, the digital production unit was working with two budgets, three grants, and seven different wage rates:

Budget #1 included wages of $8.50/hour and $9.50/hour.

Budget #2 included wages of $15.00/hour.

Work study for some students cost 25% of Budget #1 rates.

Work study for other students cost 30% of Budget #1 rates.

Grant #1 included wages of $15.00/hour.

Grant #2 included wages of $10.00/hour.

Grant #3 included wages of $8.25/hour and $8.50/hour.

DATA COLLECTION

To do a cost analysis, you must know the inputs and outputs of the project. Inputs are the time spent on various tasks, the time spent on various collections, and the labor wages. Outputs are the results of the project—in this case, the number of items added to the digital library. To quantify the inputs and outputs, you must put methods in place to collect data. As previously mentioned, the digital production unit had been collecting data since the start of the fiscal year. This data collection can be categorized into three parts: a weekly collection report, various budget and grant tracking spreadsheets, and the collection time log. The first two were managed by the digital production librarian and the latter was the responsibility of the students.

Various Budget and Grant Trackers

A different spreadsheet for each budget and each grant was kept. Some were tracked on a weekly basis; others were on a pay-period basis. The spreadsheets (figure 18.1) were structured such that only the hours worked were entered; formulas calculated amount paid per pay period, amount paid year-to-date, and actual versus budgeted dollars. This is a useful tool for projecting

work schedules to see how many students can be hired, how many hours each student can work, and how long the budgeted dollars will last. This information was used to calculate costs.

Collection Time Logs

Students received their own notebook, which remained within the digital production center. Notebooks contained a student's personal notes and other reference handouts, a log of equipment problems, and other information. The most important document in their notebook, however, is the collection time log (Pellegrino 2008). The log identifies which students are performing what tasks on which collections and for how long.

FIGURE 18.1
GRANT SPREADSHEET

Grant #2					
2008-2009		20	19	18	17
	Name	9/26/2008	10/10/2008	10/24/2008	11/7/2008
Hours	Student #28	0	0	14.5	6.5
	Student #29	12.25	16	14	15.5
	Total	12.25	16	28.5	22
Rate	Student #28	$10.00	$10.00	$10.00	$10.00
	Student #29	$10.00	$10.00	$10.00	$10.00
Expended per pay period	Student #28	$0.00	$0.00	$145.00	$65.00
	Student #29	$122.50	$160.00	$140.00	$155.00
	Total	$122.50	$160.00	$285.00	$220.00
Expended YTD	Student #28	$0.00	$0.00	$145.00	$210.00
	Student #29	$122.50	$282.50	$422.50	$577.50
	Total	$122.50	$282.50	$567.50	$787.50
Summary	Budget	$870.00	$870.00	$870.00	$870.00
	Expended YTD	$122.50	$282.50	$567.50	$787.50
	Refunded				
	Balance	$747.50	$587.50	$302.50	$82.50
	Hours YTD	12.25	28.25	56.75	78.75
	Year Remaining	77%	73%	69%	65%
	$ Remaining	86%	68%	35%	9%

Task Codes

As part of the collection time log, students entered a task code from a predetermined list (Crane and Pellegrino 2008). Task codes helped to define tasks and kept data consistent. Task codes can be pretty detailed. Think of data used for cost analysis as being captured in buckets. It is always easier to capture data in the smallest bucket and then pour this information into larger buckets, or aggregate, as needed. If data are captured at a higher level, it is not as easy, and perhaps impossible, to break up the information into more minute detail should it become necessary.

Weekly Collection Report

The digital production unit uses CONTENTdm as its digital asset management software. One feature of this software provides a snapshot of the total items included within each collection at the time the report is viewed. Each week the numbers from this report were entered into a spreadsheet. Formulas calculate the change in totals from week to week, thereby providing the incremental additions each week. These weekly additions represented output, and the total for the analysis period was used to calculate costs.

DATA ANALYSIS

This is where the sequestered number crunching session began. Information from each of the student's collection time logs was entered into an entirely new spreadsheet. It is strongly advised that the available person with the most knowledge of the digital projects do the data entry; this allows for an intimacy with the data. One becomes familiar with how each student tracks their time, what tasks are performed, and what collections each student works on. Data quality is enhanced by catching errors, or inconsistencies, in the use of task codes. After all of the information was entered, there were over 2,100 rows of data in the spreadsheet. The data were then sorted by student name, then by collection, and then by task; the results were subtotaled by collection with total hours by student (figure 18.2).

Budget Data: Hours

The next step consisted of condensing those 2,100 rows of data into a single line per student and comparing the collection time log hours with the actual paid hours taken from the various budget and grant spreadsheets. Starting with a new Excel workbook, students were listed by name down the first

column on the first sheet. Each collection was listed as a column heading across the same sheet. The total hours a student worked on a particular collection were entered into the appropriate cell (figure 18.3).

In many cases, the hours documented by the collection time log did not match the actual hours paid. Differences were then "plugged" and allocated across collections. This is where knowledge of the students and the collections they worked on came in handy. Finally, this provided total hours for each collection.

Budget Data: Wages

On a second sheet within the same Excel workbook, the students were again listed by name down the first column and each collection was listed as a column heading across the sheet. Formulas were built into each cell that

FIGURE 18.2
COLLECTED DATA

Data collected from Collection/Time Log								
Name	Collection	Task	Date	Start	Stop	Total Hrs	Completed	Hrs/Coll
student #3	cpo	ocr	12/5/2008	10.3	11.8	1.5	65	
student #3	cpo	ocr	12/10/2008	10.1	11.8	1.7	73	
student #3	cpo	ocr	12/12/2008	10.2	11	0.8	39	
student #3	cpo	ocr-cu	2/24/2009	10.1	11.8	1.7	5.5	
student #3	cpo	ocr-cu	2/26/2009	10.1	11.8	1.7	5	
student #3	cpo	ocr-cu	3/3/2009	10	11.8	1.8	6	
student #3	cpo	ocr-cu	3/10/2009	10.2	11.8	1.6	7	10.8
student #3	cwd	cp-rotate	2/17/2009	11.7	11.8	0.1	13	
student #3	cwd	me	2/19/2009	10.1	11.8	1.7	10	
student #3	cwd	sc	12/2/2008	10.5	11.8	1.3	20	
student #3	cwd	sc	2/12/2009	10.1	11.8	1.7	1	
student #3	cwd	sc	2/17/2009	10.2	11.4	1.2	22	
student #3	cwd	sc/cp	10/2/2008	10.15	11.55	1.4	14	
student #3	cwd	sc/cp	10/7/2008	10.1	11.55	1.45	22	
student #3	cwd	sc/cp	10/9/2008	10.35	12	1.65	28	
student #3	cwd	sc/cp	10/14/2008	10.3	11.55	1.25	??	
student #3	cwd	sc/cp	10/16/2008	10.3	11.55	1.25	22	
student #3	cwd	sc/cp	10/23/2008	10.1	11.3	1.2	26	
student #3	cwd	sc/cp	10/28/2008	10.1	11.5	1.4	23	
student #3	cwd	ts	12/12/2008	11.1	11.8	0.7	??	16.3
student #3	sca	me	12/2/2008	10.1	10.3	0.2	26	
student #3	sca	me/mn	11/4/2008	10.2	11.45	1.25	3	
student #3	sca	me/mn	11/6/2008	10.3	11.45	1.15	7	
student #3	sca	me/mn	11/11/2008	10.3	11.5	1.2	8	
student #3	sca	me/mn	11/18/2008	10.15	11.45	1.3	5	
student #3	sca	me/mn	11/20/2008	10.1	11.45	1.35	15	
student #3	sca	measuring	11/20/2008	10.1	11.15	1.05	31	7.5
						34.6		34.6

FIGURE 18.3
BUDGET DATA: HOURS

CCDL Costing March 2009
Budget Data - Hours

Collection alias →

Student Name	Total hours as of 02/27/2009	ck	aot	bce	bia	ccp	cdl	cmt	cpo	csa	cwd	dac-doc	da[illegible]	[illegible]m	nha	pap	per	phl	sca	twi	wsc-file	wsc-ocr	non prod	Total per Ntbk	Total per Other	Total Hrs	Total Hrs v. Budget
Student #1	278	0				19	5				5													247	30		
Student #3	36	0							11		16								9					35	1		
Student #7	69	0														5	65								69		
Student #8	150	0									127			3					8					155	-5		
Student #9	55	0									55													46	9		
Student #10	117	0									81								6				2	127	-11		
Student #12	110	0									30							9			18	19	2	77	33		
Student #22	29	0									29													43	-14		
Student #23	20	0	5								15													20	-1		
Budget #1	**2,172**																							**1,372**	**800**	**2,172**	0.00
Student #24	150	0		150																					150		
Student #25	64	0				64																			64		
Student #26	88	0				88																			88		
Student #27	85	0				85																			85		
Grant #1	**387**																							**0**	**387**	**387**	0.00
Student #28	21	0																							21		
Student #29	58	0																						22	35		
Grant #2	**79**																							**22**	**56**	**79**	0.00
Student #30	33	0																						43	-10		
Student #31	22	0																							22		
Student #32	66	0																							66		
Student #33	12	0																						19	-7		
Grant #3	**132**																							**62**	**70**	**132**	0.00
Student #34	386							368				18												386			
Student #35	676							555																676			
Budget #2	**1,062**																							**1,062**	**0**	**1,062**	0.00
Totals - Prod & Non-prod	**3,833**		5	150	6	309	5	924	11	2	675	18		3	74	5	65	92	26	1	26	19	136	2,519	1,314	3,833	
Less - Non-production	-136																					3697	3832				
Total - Production only	3,697																										

FIGURE 18.4
BUDGET DATA: WAGES

CCDL Costing March 2009
Budget Data - $$

Student Name	Total hours as of 02/27/2009	Rate	%	Total spent	Collection alias → aot	bce	bia	ccp	cdl	[illegible]	nm	nha	pap	per	non production	Total sum of coll	Rnding diff
Student #1	278	$9.50	100%	$2,636.25	$0.00	$0.00	$0.00	$178.13	$51.21	$0	$0.00	$0.00	$0.00	$0.00	$0.00	$2,636.35	0.10
Student #3	36	$9.50	100%	$339.15	$0.00	$0.00	$0.00	$0.00	$0.00	$0	$0.00	$0.00	$0.00	$0.00	$0.00	$339.15	0.00
Student #7	69	$8.50	100%	$586.50	$0.00	$0.00	$0.00	$0.00	$0.00	$0	$0.00	$0.00	$38.25	$548.25	$0.00	$586.50	0.00
Student #8	150	$8.50	30%	$382.50	$0.00	$0.00	$0.00	$0.00	$0.00	$0	7.65	$0.00	$0.00	$0.00	$0.00	$382.50	0.00
Student #9	55	$8.50	25%	$117.58	$0.00	$0.00	$0.00	$0.00	$0.00	$0	$0.00	$0.00	$0.00	$0.00	$0.00	$117.62	0.04
Student #10	117	$8.50	25%	$248.26	$0.00	$0.00	$0.00	$0.00	$0.00	$0	$0.00	$0.00	$0.00	$0.00	$4.25	$248.20	-0.06
Student #12	110	$8.50	100%	$935.85	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$17.00	$935.85	0.00
Student #22	29	$8.50	100%	$246.50	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$246.50	0.00
Student #23	20	$8.50	100%	$165.75	$42.50	$0.00	$0.00	$0.00	$0.00	$0	$0.00	$0.00	$0.00	$0.00	$0.00	$165.75	0.00
Budget #1	**2,172**			**$16,468.35**												**$16,468.41**	0.05
Student #24	150	$15.00	100%	$2,250.00	$0.00	$2,250.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$2,250.00	0.00
Student #25	64	$15.00	100%	$960.00	$0.00	$0.00	$0.00	$960.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$960.00	0.00
Student #26	88	$15.00	100%	$1,320.00	$0.00	$0.00	$0.00	$1,320.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$1,320.00	0.00
Student #27	85	$15.00	100%	$1,275.00	$0.00	$0.00	$0.00	$1,275.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$1,275.00	0.00
Grant #1	**387**			**$5,805.00**												**$5,805.00**	0.00
Student #28	21	$10.00	100%	$210.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$210.00	0.00
Student #29	58	$10.00	100%	$577.50	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$577.50	0.00
Grant #2	**79**			**$787.50**												**$787.50**	0.00
Student #30	33	$8.25	100%	$268.95	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$268.95	0.00
Student #31	22	$8.25	100%	$181.50	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$181.50	0.00
Student #32	66	$8.25	100%	$542.44	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$542.44	0.00
Student #33	12	$8.50	100%	$102.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0	.00	$0.00	$0.00	$0.00	$0.00	$102.00	0.00
Grant #3	**132**	$8.31		**$1,094.89**												**$1,094.89**	0.00
Student #34	386	$15.00	100%	$5,790.75	$0.00	$0.00	$0.00	$0.00	$0.00	$5,52	.00	$0.00	$0.00	$0.00	$0.00	$5,790.75	0.00
Student #35	676	$15.00	100%	$10,144.65	$0.00	$0.00	$0.00	$0.00	$0.00	$8,33	$0.00	$0.00	$0.00	$0.00	$0.00	$10,144.65	0.00
Budget #2	**1,062**			**$15,935.40**												**$15,935.40**	0.00
Totals - Prod & Non-Prod	3,833			**$40,091.14**	$42.50	$2,250.00	$53.20	$4,241.38	$51.21	$13,85	7.65	$698.25	$38.25	$548.25	$1,273.35	$40,091.19	0.05
Less - Non-production				**-$1,273.35**													
Total - Production only				**$38,817.79**													

multiplied the hours from the first sheet ("Budget data—hours") with the hourly rate and percentage. This resulted in a total cost per collection at the bottom of the sheet (figure 18.4).

RESULTS

For each collection, the following were identified as part of this exercise:

- total items added over the fiscal year (taken from the weekly collection report)
- total cost
- average cost per item
- average rate per hour
- total student hours
- average items per hour
- average minutes per item

Of course, some information could not be determined. In some cases, the student labor and hours were tracked by another department and were not accessible for this exercise. In other cases, only digital production staff worked on the collection. Since they were not part of this exercise, these calculations were not done. But, for the most part, considerable information was identified (figure 18.5).

Data Correction and Validation

Some of the results seemed to be on the high side; they stood out and warranted further investigation. In a review of the data it became apparent that the total number of items added to a collection was too low for some of our more complex items, such as monographs. For example, forty-three monographs containing a total of 766 chapters were added to a single collection at a cost of $3,393. Because each chapter was digitized as a PDF and each monograph required descriptive metadata and digital assembly, a denominator of 809 (43 monographs + 766 chapters) was used, resulting in a cost per item of $4.19, which is too high. Once it was realized that the time it took to scan each page of each chapter was not taken into consideration, a new denominator of 7,025 (43 monographs + 766 chapters + 6,216 pages) was used, resulting in a more reasonable $0.48 cost per item. The $3,000-plus dollars represented the time and cost for scanning each page, running the scans through optical character recognition software, creating the PDF, uploading into CONTENTdm, and creating metadata for the single compound object, so

FIGURE 18.5
RESULTS OF THE ANALYSES

Digital Production Services
Costing
March 2009
Collection Data

Collection Name	Total new items added as of 02/26/2009	Total used for calculations	Note	Total Cost	Average Cost per Item	Average Rate per Hour	Student Production Hours	Average Items per Hour	Average Minutes per Item
Antiquities of the Institute for Antiquity and Christianity	533 (97 cpd/ 436 images)	50	Descriptive metadata only	$42.50	$0.85	$8.50	5.0	10	6
Bulletin of the Institute for Antiquity and Christianity	22 (1 cpd/ 21 pdf/92 pgs)	92	Scanning/full metadata	$53.20	$0.58	$9.50	5.6	16	4
Boynton Collection of Early Claremont	232	232	Scanning/full metadata	$2,250.00	$9.70	$15.00	150.0	2	39
California Water Documents	829 (43 cpd/766 pdf/6,216 pgs)	7,025	Scanning/PDF/OCR/ full metadata	$3,393.12	$0.48	$5.03	674.7	10	6
Campi Phlegraei	111	11	OCR & OCR clean-up	$102.60	$9.33	$9.50	10.8	1	59
Chikanobu and Yoshitoshi Woodblock Prints	36	---	CCDL staff work	---	---	---	0.0	---	---
Claremont Colleges Photo Archive	980	980	Scanning/full metadata	$4,241.38	$4.33	$13.72	309.3	3	19
Claremont Colleges Sustainability Archive	38 (5 cpd/33 pdf)	5	Full metadata only	$15.20	$3.04	$9.50	1.6	3	19
Claremont Discourse Lectures	3	3	Videos	$51.21	$17.07	$9.50	5.4	1	108
Connie Martinson Talks Books	300	300	Videos	$13,854.90	$46.18	$15.00	923.7	0	185
Drucker Archives	172	151	Scanning/full metadata	$267.00	$1.77	$15.00	17.8	8	7
	---	21	Videos	$1,813.50	$86.36	$15.00	120.9	0	345
Early California Letters	621	---	Special Collections	---	---	---	---	---	---
Faculty Scholarship at The Claremont Colleges	55	55	U-SKIs prep work/metadata	$4,023.08	$73.15	$9.49	423.9	0	462
Fashion Plate Collection, 19th Century	222	222	Archival processing/ scanning/full metadata	$1,484.25	$6.69	$9.55	155.5	1	42
Growing a Digital Library	2	---	CCDL staff work	---	---	---	---	---	---
History of Jazz 109	14	14	Full metadata only	$43.70	$3.12	$9.50	4.6	3	20
Interface Journal Archive	1 (1 pdf/24 pgs)	24	Scanning/full metadata	$11.40	$0.48	$9.50	1.2	20	3
Italian Renaissance Manuscript Collection	53	---	Special Collections	---	---	---	---	---	---
Ken Gonzales-Day Collection	181 (161 items/ 18 pdf/437 scans)	616	Scanning/full metadata	$1,067.99	$1.73	$7.39	144.5	4	14
Larry Oglesby Collection	1,362	1,362	Scanning/full metadata	$1,669.73	$1.00	$8.66	192.9	7	8

the number of pages scanned had to be included in the divisor. This required a review of all the underlying denominators for each collection and resulted in a "total used for calculations" (see figures 18.4 and 18.5) that went beyond the original denominator of new items added to a collection.

Formulas

A variety of formulas were used to calculate each of the items identified above:

Total cost was derived from the bottom of the "Budget data: wages" sheet by collection.

Student production hours were derived from the "Budget data: hours" sheet by collection.

Average cost per item = Total cost ÷ total used for calculations.

Average rate per hour = total cost ÷ student production hours.

Average items per hour = total used for calculations ÷ student production hours.

Average minutes per item = (student production hours ÷ total used for calculations) x 60.

Additional Cost Factors

This exercise focused solely on student wages because their wages were considered direct costs of digital projects. If there were no digital projects, there would be no student costs. To do a complete cost analysis of digital projects, however, one must also take into account indirect costs, such as direct and indirect staff wages, hardware and software maintenance costs depreciation, and other overhead such as allocations for square footage and utilities.

CONCLUSION

After completing this project, the digital production librarian was able to answer the original question—"How much does it cost to put an item into the digital library?"—with "It depends." Because of the variety of material types and the range of funding sources and wage rates that characterize the inputs to the CCDL, it was difficult to provide an uncomplicated answer.

Once the results were in, many additional uses for this information became apparent. With some supplementary calculations, a cost by material type was determined and the time and cost for scanning and metadata creation for a

particular collection were separately identified. Those libraries with projects utilizing fewer material types or a smaller wage variance should be able to derive a comprehensive result.

Managers who understand and quantify the inputs and outputs of a digital project and use some of the tools and methods presented in this case study have a place to start costing their own projects should their administration or external funding sources ask, "What do your digital projects cost?"

WORKS CITED

Crane, Lisa L., and S. Pellegrino. 2008. Task codes. http://ccdl.libraries.claremont.edu/u?/adl,48 (accessed February 26, 2010).

Pellegrino, Sharon. 2008. Collection time log. http://ccdl.libraries.claremont.edu/u?/adl,49 (accessed February 26, 2010).