An Experiment to Increase Online Archival Accessibility: Using Unique Page Views to Measure Online Efficiency

Jonathan Dembo and Mark Custer

Background

Over the years, archivists, librarians, manuscript and special collection curators and their staffs and repositories[1] have responded to the changing needs of their administrators, accreditation bodies, and professional colleagues and, adapting to rapidly changing technologies, have compiled a wealth of statistical data on their activities. In the view of the authors, their application of this data in their daily work, however, has failed to keep pace. Today, most major non-archival organizations far surpass archival repositories in their use of available statistics for self-analysis. Long before a major organization introduces a new product, for example, it tests its various qualities through surveys, focus groups, and market studies. Sometimes it produces several possible products and tests each against the other and their major competitors before deciding to mass-produce or cancel a production project. It then quantifies these results and compares them with competing products made by themselves and other corporations with various qualities and prices. Millions or even billions of dollars may ride on their decisions. This kind of testing continues through production and marketing phases and continues all during the life cycle of the product. The authors believe that archivists can and should do something similar, albeit at a lower level of expense, to assess their online finding aids.

Repository staffs do not lack for available statistical data with which to test their online finding aids. Archivists count the numbers of users both online and in research rooms, sometimes by time of day. They count the numbers and kinds of reference and research questions they receive and whether received in person, by letter, by telephone, by email, or by web form. They count the number of requests by collection, by subject, by sex, age, and address or zip code of researcher, by staff member responding, and by a myriad of other categories.[2] The authors think it fair to say that despite this demonstrated interest in compiling statistics of various kinds, archivists can and should make much better use of this information.

Theses

The authors believe, for example, that at present most archivists and manuscript curators cannot say with any degree of certainty why researchers access their online finding aids or use their collections, or prefer one collection over another. Are researchers using some collections because they are interesting or useful or simply well known, or all three, or none of the above? Are researchers using some collections because of their content or because of how archivists have arranged or described them? Are researchers using some collections because the online finding aids are usable and informative or both or neither?[3] In other words, are researchers using collections because of something inherent in the collections, or because of something caused by archivists?

The authors further believe that archivists and manuscript curators may use easily available online access statistics to compare the effectiveness of their online finding aids, from most to least accessed; that they may use experimental changes to their least accessed online finding aids to determine why these finding aids have not attracted more researchers; and that they may use experimental changes to specific components of their least accessed online finding aids to identify those components most in need of revision. In so doing, the authors believe that archivists may acquire an effective tool to improve the online accessibility of all their finding aids.

The Experiment

The authors have therefore conducted the present experiment, which they believe any repository may replicate. The results can enable a repository staff to use their own online access statistics to analyze the effectiveness of their online finding aids and, hence, to develop ways to improve their internal descriptive procedures and to increase public access to their collections. The authors believe that this technique is adaptable to any repository whether or not its finding aids are available online and may be used to test

the effectiveness of many different archival functions, including processing, arrangement, description, and cataloging. They also feel that repositories may also apply the technique to improving their finding aid and webpage design and navigation.

Joyner Library had been recording and analyzing online user statistics for a number of years. A grant project in 2001-2003, funded by NC ECHO[4] and the North Carolina State Library, digitized and encoded the finding aids to the Special Collections Department's manuscript collections using Encoded Archival Description.[5] By 2008, the finding aids to virtually all of Joyner Library's manuscript collections were available online.[6] By that time, too, Joyner Library had begun to use Google Analytics to record online usage of its website. The data from Google Analytics allowed the authors to easily track usage of individual online finding aids on a daily, weekly, monthly or cumulative basis for the first time.[7]

Joyner Library's project began in August 2009, when Special Collections Curator Jonathan Dembo began to examine the number of Unique Pageviews (UPVs) obtained by online viewers who had examined special collections online finding aids during the previous year (2008/2009). He had been tasked to find ways to revise the system the Special Collections Department used to determine the subject terms added to online finding aids. In the course of his research, he found that some online finding aids received many UPVs month after month; many more received a few such UPVs; but most online finding aids received no UPVs during an average month. There did not seem to be any reason why some finding aids received more UPVs than others, except that the most-used collections tended to have more elaborate and detailed finding aids and the less-used collection tended to have less elaborate and detailed finding aids. It seemed clear to the authors that there was no direct relationship between the importance of the collection and the number of UPVs the online finding aid received. It seemed equally

obvious to the authors that the quality of the online descriptions of the collections had something to do with the online use of the finding aids. However, it was entirely unclear in what way the finding aids were affecting the outcome. The question was: how to use the available information to show how the finding aids were impacting the statistical results?

The Hypothesis

The authors reasoned that they had three possible ways to use the statistics to assess the quality of Joyner Library's online finding aids. First, they could reduce the quality of the information in the most frequently accessed guides to see how that affected usage of the most popular guides. Second, they could try to improve the quality of these most frequently accessed finding aids to see if this increased usage. Alternatively, they could try to improve the quality of a selection of the least used guides to see if this would influence their usage.

The authors immediately discarded the first two alternatives. It seemed to them irresponsible to diminish the quality or quantity of description available to online researchers for any collection. It also seemed to them that it would be too much work to try to improve the already large and high quality online finding aids available for many of the already popular collections. The authors reasoned, instead, that they should concentrate on improving the quality and amount of information available for a selection of the least used online finding aids. They reasoned that the guides to these collections would be both smaller and easier to improve than if they chose a selection of the most accessed online finding aids. A similar group of least used finding aids would be selected as a control sample.

The authors reasoned that if they improved the least used online finding aids and the number of UPVs did not increase more than the control sample, it would tend to prove that the collections lacked importance. If, on the other hand, they improved the finding aids and the number of UPVs increased more than the control sample, it would tend to prove that the finding aids themselves were responsible rather than the importance of the collection. By selectively revising different components of the online finding aids and tracking the subsequent UPV statistics, the authors also hoped to shed light on those specific elements of the finding aids that had

the greatest and least impact on the overall results. Methodology: In order to test this hypothesis that improvements in the online finding aids could impact online statistics, Prof. Dembo obtained the assistance of Joyner Library Digital Collections faculty member, Prof. Mark Custer, whose responsibilities included compiling the online statistics for Joyner Library. On 22 July 2009, Prof. Custer prepared a list including all the online finding aids available during the entire 12-month period ending 30 June 2009.[8] The Special Collections EAD website produced 8,527 pages during that period; these pages received a total of 83,387 UPVs during the preceding 12-month period. Of the 1,762 online finding aids on this list, 114 finding aids had received four or fewer (later adjusted to six or fewer) UPVs during the year sorted according to the number of UPVs that each online finding aid received.[9]

From this list of least accessed online finding aids, Prof. Dembo simply selected the first five and last five online finding aids on the list that described collections containing less than one and a half cubic feet of manuscript materials regardless of processing status. He excluded non-manuscript collections such as maps and oral histories. The first five selected guides each received four to six UPVs during the year ending 30 June 2009; the last five had each received no more than one UPV during the same year.[10] He next selected the five online guides on the list immediately adjacent to each group as a control sample. Upon examination, he found that all the online guides selected represented collections he had hitherto thought unimportant or uninteresting because researchers rarely tried to use them. Most were virtually unprocessed. They had brief and non-detailed finding aids.

Table 1: Revised Finding Aids Statistics: Test Sample, July 2007 – Oct. 2010, lists the ten tested online finding aids in the experiment and records the number of UPVs obtained during the test period. Prof. Dembo then selected the adjacent five online finding aids from each group to serve as a control sample. Table 2: Revised Finding Aid Statistics: Control Sample, July 2007 – Oct. 2010, lists the ten online finding aids in the control sample and records the number of UPVs obtained during the test period.[11]

Altogether, the authors included twenty online finding aids in their study. The finding

aids in the tested and control samples appeared to be very similar, generally. The collections in the tested sample were somewhat larger on average than the control sample. The tested sample averaged 2.703 cubic feet; the control sample averaged 1.723 cubic feet.[12]

During 23-24 July 2009, Prof. Dembo revised the first group of five online finding aids. He enhanced or added descriptive information and historical detail, including biographical and historical notes, collection inventories, scope notes, and accession information to the online finding aids. He did not rearrange or reprocess the collections themselves.[13]

Table 3: Changes to Online Finding Aids, 2009-2010 indicates how many words Prof. Dembo added to each section of each of the online finding aids. Table 3 records the collection names and numbers for each of the tested finding aids. It also records the number of words added to the finding aid and indicates whether the words were added to the Biographical / Historical Notes, the Inventory, the Scope Note, or the Accession Information.[14]

Table 4: Ranking Changes to Tested Finding Aids, July 2007 – June 2010, shows how the finding aids changed relative position from 2008-2009 to 2009-2010. In 2008-2009 the ten tested finding aids average ranking was 2,368.3 out of a total of 8,527 pages on the website; by 2009-2010, after the experiment, they ranked 929.7 out of a total of 12,492 pages on the website. They had an average increase of 1,438.6 positions per finding aid. All this for an average increase of only 850.6 words per finding aid.

Prof. Mark Custer posted the first five revised finding aids on the Joyner Library website during 10-12 August 2009 so the figures for August 2009 cover only part of the month. Due to time constraints, Prof. Dembo was not able to complete work on the second five collections until 12 February 2010. Prof. Custer posted the second five revised finding aids 12-15 February 2010 so the figures for February 2010 cover only part of that month. Thus, during the test period, the first five revised finding aids were online from mid-August 2009 through October 2010; the second five revised finding aids were online only from mid-February 2010 through October 2010, or five months less than the first group.

By the end of December 2009, however, the first group of tested finding aids all showed a dramatic rise in UPVs. By the end of December 2009, after only four and an half months, the first five tested online finding aids had received 99 UPVs, three times more than all ten online finding aids had received in all 12 months of 2008-2009. This was far in excess of the number of UPVs received by the control sample. All ten online finding aids in the control sample had received only 30 UPVs which was less than one third the totals received by the tested sample.[15]

This ratio persisted over the next four months. Table 1 shows that by April 2010, the number of UPVs per tested collection increase ranged from 12 to 22. In the year prior to the experiment, the ten tested finding aids received a total of 30 UPVs or an average of 3 UPVs per finding aid per year. In the first year after the experiment began, the ten tested finding aids received a total of 314 UPVs or an average of 26.2 UPVs per month, or more than 31 UPVs per finding aid per year. This 1,046.66%, increase was achieved despite the fact that the first five tested finding aids were online for only 10.5 months and the second five tested finding aids were only online for 4.5 months.[16] In FY 2010/2011 a similar pattern has continued at a somewhat reduced level. During July – Oct 2010, the ten tested finding aids received 91 UPVs or an average of 22.75 UPVs per month. If the same rate persists for the rest of FY 2010/2011, the tested sample will receive 273 UPVs, or more than 27 UPVs per finding aid per year.

Table 1 also shows that prior to the experiment the tested sample received an average of 2.5 UPVs per month in FY 2008/2009; after revision, the tested sample received an average of 26.17 UPVs per month in FY 2009/2010. During FY 2009/2010, the tested sample's UPVs jumped from 30 to 314. During the 16-month period of the experiment, the tested sample received an average of 25.31 UPVs per month a 1,046.66% increase over FY 2008/2009.

Table 2 shows the results for the control sample over the same period of time. For the period the control sample behaved far more typically of the website as a whole than the tested finding aids. Table 2 shows that the control sample received 30 UPVs or an average of 2.5 UPVs per month in FY 2008/2009. In the year of the experiment, the control sample

UPVs rose to 57 or an average of 4.75 UPVs per month or a 90% increase. For the entire 16 month period of the experiment, the control sample received a total of 90 UPVs, or an average of 5.63 UPVs per month. This represents an increase of 225.2% over FY 2008/2009, but was far below the increase gained by the tested sample.

Moreover, readers should note that the statistics are somewhat skewed in favor of the control sample by one finding aid. More than half of the total UPVs (17 of 30 UPVs) in FY 2009/2010 were received by this single finding aid during a single month.[17] And all the control sample finding aids were online for the entire period, whereas the tested sample were online for significantly less time.

During the July – Oct. quarter of FY 2010/2011, the ten control finding aids obtained 33 UPVs or an average of 3.3 UPVs per finding aid. If the same rate persists for the rest of FY 2010/2011, the control sample will receive 99 UPVs for the year, or only about 3.3 UPVs per finding aid.[18] In other words, the control sample began at virtually the same level as the tested sample and only increased about 10% during the test, which is below the general increase in UPVs for the website as a whole.[19]

The authors next considered whether changes to particular elements of the finding aids had particular effects. Table 3: Changes to Online Finding Aids, 2009-2010 shows the number of words Prof. Dembo added to the finding aids in the tested online finding aids in an attempt to answer this question. Overall, Prof. Dembo added an average of 851 words per online finding aid tested. He added biographical and or historical notes to four of the finding aids; he added a preliminary inventory to nine of the finding aids; he added scope notes to three of the finding aids; and he added accession information to four of the finding aids.[20] He added an average of 591 words to the first five finding aids tested; he added an average of 1110.2 words to the second five finding aids tested.

Table 4: Ranking Changes to Finding Aids, July 2009 – June 2010, compares the relative ranking for each of the finding aids in the experiment both before and after the experiment. It includes the number of words added to each finding aid during the experiment and the number of ranking places

each finding aid changed during the collection. Table 4 shows that during FY 2008-2009, the total number UPVs received by the entire website stood at 83,387, or an average of 9.8 UPVs per page. During FY 2009/2010, the total number of UPVs received by the website increased 4.43% from 83,387 to 87,084, but this represented a decline to an average of 7 UPVs per page. Meanwhile, the tested sample went from a below average 3 UPVs per month to 25 UPVs per month, in FY 2009/2010, nearly triple the website average.[21]

In addition, Table 4: Ranking Changes to Tested Finding Aids, July 2009 – June 2010, shows how the experiment affected the relative rankings of the online finding aids.[22] In 2008-2009, the year prior to the experiment, online researchers accessed a total of 8,527 separate finding aid files amounting to 83,387 for the website as a whole. The average rank of the ten online finding aids tested was 2,368 out of 8,527, or in the 27.5 percentile of online pages. The ten finding aids tested ranked near the bottom of the list of finding aids with at least 1 UPV during the previous year.

Table 4 also shows that in FY 2009/2010, the year of the experiment, the number of finding aids on the website had increased to 12,492, a 46.5% increase over the previous year. However, online access statistics did not keep pace. Online researchers made a total of 87,084 UPVs, a rise of only 4.25% over the year before. This time, however, the finding aids used in the collection had an average rank 929.7 out of 12,492, which was in the 7.45 percentile of the online web pages, which was a dramatic improvement. Despite a nearly 50% increase in the number of web pages during the year, the tested sample improved its relative standing from 1727 to 929.7. In other words, the tested finding aids shot up an average of 1,438.6 rank places after having an average of 850.6 words added to each and the overall tested sample raised an average of 1.69 rank places for each word added to the finding aid. However, this left unanswered the more important questions of whether adding text to specific elements of the online finding aids would have been more or less effective in raising online usage of the finding aids.

The number of words added seemed to be independent of the results recorded. For example, in FY 2008/2009, Democratic Women of North Carolina Collection (#518) was the highest ranked finding aid at position 1,484.

During the experiment, Prof. Dembo added 157 words to the online finding aid. He added 110 words to the preliminary inventory and 47 words to the accession notes. Overall, this was the second smallest number of words he added to any of the finding aids. Nevertheless, while Collection #518 rose 772 ranking places, it fell to fourth place among the ten tested collections in FY 2009/2010. Additionally, this was the smallest rank increase among all the online finding aids.[23]

On the other hand, the John Vainwright Bible Records (#MG0063) was the tenth ranked online finding aid in FY 2008/2009 at position 3,072. During the experiment, Prof. Dembo added 250 words – the third smallest number of words added to the online finding aids. He added 127 words to the preliminary inventory, 88 words to the scope notes, and 35 words to the accession notes. Unexpectedly, Collection #MG0063 rose 1,399 ranking places in FY 2009/2010. It rose from last place to fourth place at rank 1,673.[24]

Initially, the Clio Book Club Records (#579) was the eighth ranked online finding aid in FY 2008/2009 at position 3,055. During the experiment, Prof. Dembo added 2,217 words to the online finding aids, the most words added to a finding aid. He added 2,163 words to the preliminary inventory and 54 words to the accession notes. As a result, collection #579 rose 1,935 places in FY 2009/2010 to 1,120th place. Nevertheless, it remained in eighth place among the tested finding aids.

On the other hand, the Martha E. Donaldson Papers (#517) was the fifth ranked online finding aid in FY 2008/2009 at position 1,732. During the experiment, Prof. Dembo added 772 words to the online finding aids. He added 220 words to the biographical / historical notes, 512 words to the preliminary inventory, and 40 words to the accession notes. This was the fifth highest total of words added to any of the finding aids. As a result, Collection #517 rose 1,323 ranking places in FY 2009/2010 to become the highest ranking online finding aid tested at rank 409.

Prof. Dembo, therefore, compared the impact on the finding aids of words added to the various elements tested: Biographical/Historical Notes, Inventories, Scope Notes, and Accession Information. Table 5: Impact of Changes to Finding Aid Elements, 2009-2010 shows that the number of words added to

the various parts of the tested finding aids did have a direct, consistent, impact on the relative rankings of the finding aids tested.[25] Prof. Dembo added an average of 457.75 words to the Biographical/Historical Notes elements of four of the tested online finding aids. In 2008-2009 these had an average rank of 2,059.25 out of 8,527 online pages. In 2009-2010 they had an average rank of 611.25 out of 12,492, which represented a rise of 1,447.5 ranking places. These collections improved an average number of 5.83 ranking places per word added.

Prof. Dembo added an average of 690.33 words to the Inventories elements of nine of the tested online finding aids. In 2008-2009 these had an average rank of 2,439.44 out of 8,527 online pages. In 2009-2010 they had an average rank of 961.44 out of 12,492, which represented a rise of 1,478 ranking places. These collections improved an average number of 6.15 ranking places per word added.

Prof. Dembo then added an average of 89.33 words to the Scope Note elements of three of the tested online finding aids. In 2008-2009 these had an average rank of 2,616.66 out of 8,527 online pages. In 2009-2010 they had an average rank of 852.66 out of 12,492, which represented a rise of 1,763 ranking places. These collections improved an average number of 22.44 ranking places per word added.

Finally, Prof. Dembo added an average of 89.33 words to the Accession Information elements of four of the tested online finding aids. In 2008-2009 these had an average rank of 2,664.5 out of 8,527 online pages. In 2009-2010 they had an average rank of 1,325.5 out of 12,492, which represented a rise of 1,339 ranking places. These collections improved an average number of 28.45 ranking places per word added.

As a result, it would seem that based on the online finding aids tested, archivists at Joyner Library should place their priority on providing online users with enhanced information in the following areas:

1. Accession information (28.45 average rank improvements per word added).
2. Scope Note (22.44 average rank improvements per word added).
3. Inventories (6.15 average rank improvements per word added).
4. Biographical / Historical Note (5.83 average rank improvements per word added).

Conclusions

The authors would like to caution, however, that their conclusions are tied to how the finding aids at Joyner Library are currently presented online. Since our "Accession information" occurs within the first HTML paragraph tags on those webpages, it is likely that those terms are indexed by external search engines with an increased weight over those terms that appear near the end of our HTML finding aids, such as those within the "Inventories."

Moreover, the authors also desire to note that the present experiment was a trial run based on a very small sample. Many other variables – including the quality of the specific changes made to the finding aids – may have played important but unappreciated roles in the experiment. The authors are convinced, however, that the revisions of the tested online finding aids did have a decisive impact in increasing the number of UPVs for the tested finding aids received as a group. The control sample in the test showed very little change in comparison.

The authors are unaware of previous attempts to use online access statistics to identify and select archival finding aids for experimental purposes. They are also unaware of any previous attempt to use online access statistics, Unique Page Views (UPVs) or Google Analytics to track and assess the results of finding aid revision experiments. The authors are convinced, however, that archivists may use this technique with profit to explore the effectiveness of their own online finding aids and to explore how and where to concentrate their efforts to improve them.

The authors have shown that, for very little expense, time, and labor (less than two working days overall) they have at least tripled the UPVs obtained by ten of Joyner Library's online manuscript finding aids. They have accomplished this while the total number of online pages rose by less than 32% and the total number of UPVs increased by only 4.25% during the test period.

The authors are also convinced that their technique may allow archivists to increase online access to their collection finding aids without necessarily having to acquire, process, or catalog entirely new collections and without having to actually process or re-process the collections themselves.[26]

The authors believe that they have also provided strong evidence that changes in particular elements of the online finding aids tested can have even greater and measureable results. Their attempt to demonstrate that archivists may be able to increase the effectiveness of their own online finding aids by making changes to particular sections of these finding aids and tracking the results over time showed that improvements to the Scope Notes and Accession Information elements of Joyner Library's online finding aids had a disproportionate beneficial impact in generating UPVs and raising the rank of the finding aids changed. It also showed that changes to the Biographical/Historical Notes and Inventory sections had a less dramatic impact in generating UPVs and raising the rank of the finding aids changed.

The authors suggest that this study be taken as a model for future, larger and longer term studies, both in Joyner Library and elsewhere. Future samples should include a greater variety of online finding aids. Future samples should also include a wider variety of online finding aid elements. If future studies can confirm the findings produced in this experiment, they may provide archivists with an invaluable tool to make their online finding aids much more useful to researchers.

Notes on Technical and Statistical Methods

Joyner Library originally outsourced the creation of its EAD finding aids in 2002 and has been collecting Web log analysis information about its online finding aids by a variety of methods ever since. This study, however, only addresses the web metrics that have been gathered with our Google Analytics account, which we first started collecting data with on June 9, 2008.

Within Google Analytics, there are a variety of predefined metrics that are calculated. These metrics include number of visits, referring sites, new vs. returning visitors, and much more. For the purposes of our data analysis, the authors have decided to only examine the Unique Page Views metric (UPVs), which is a subset value of the Page Views metric.[27]

UPVs were chosen because this particular metric is most closely aligned with how we gather collection usage statistics within our physical reading room. Within Google Analytics, UPVs are defined as an aggregation of Pageviews that are generated by the same user, during the same session. A single user, according to how Google Analytics collects data, can more accurately be defined as a single web browser. The session, then, is comprised of the length of time that a specific web browser interacts with a website. If thirty minutes of inactivity occur, however, then that particular session will be terminated.[28] Therefore, if a user would happen to visit a single page on our website ten different times within a ten minute timeframe, only a single UPV would be registered.[29] Somewhat similarly, if a user were to request to look at material within our physical reading room, we would track that usage by their initial request per day, not by their subsequent level of interaction with the material.

However, none of the metrics gathered by Google Analytics, including the UPVs, should be considered to fully correspond to any physical reading room statistics. For instance, Google Analytics will not collect data if users have JavaScript disabled from within their web browser or if they are accessing the information in a way that would not (or could not) initiate the JavaScript code. Further, if users prevent their browsers from accepting first-party cookies, Google Analytics will not track any information. Alternatively, users may choose to delete cookies before their specific expiration dates, thus changing how Google Analytics collects the metrics for different users.[30] Nevertheless, this study is only interested in comparing how UPVs change (or do not change) over time, and it therefore should not matter exactly how every piece of (potential) data is gathered by Google Analytics or not.

For future studies conducted, though, it might be advisable to eliminate specific sets of internal traffic. This particular study intentionally did not make attempts to filter out any data that was collected during the process. We were not concerned with temporary spikes in our data.[31] Our major concern, rather, was that our data was collected consistently in order to track any trends regarding the usage of each online finding aid that was enhanced for the project. Admittedly, these collections might initially see a small artificial increase in usage based solely on staff examinations of the finding aids (i.e., internal traffic). However, it is not expected that this increased usage would be sustained over time. Therefore, if any of the collections do maintain a sustained increase in usage, the authors believe that such an increase should be the result of the revisions alone. Nevertheless, it is possible for any additional studies to exclude data that is collected. In fact, Google Analytics accommodates this by incorporating a predefined filter which can exclude all traffic from specific Internet Protocol addresses.[32]

For more detailed and current documentation about Google Analytics, it is best to visit their website directly at http://code.google.com/apis/analytics/.

Tip for Exporting Records from Google Analytics

If your site has a lot of content, you may find it difficult to export your records initially. By default, you can only export your data in the range of 10, 20, 100, or 500 rows at a time. Considering that you might want to export a much larger batch of data than this, the export process could quickly become tedious if you did not implement an alternative method. Even without taking advantage of the Google Analytics API, a very simple method exists to increase your export limit. To employ this method, first check to see how many rows of data that you have for any given report. For instance, one of our "Top Content" reports for a single month is listed as having "2,218" rows of data. Look in the lower right-hand corner of your table for this value:

Rather than increase the visible rows to 500, and exporting 5 different data sets, all that needs to be done is to add a special limit to the end of the URL:

&limit=2218

...where the number after the "=" sign is >= to the number of rows of data that your particular report contains.

This will not increase the amount of records that you can view from within the Google Analytics interface at any one time, but it will permit you to export all of those records into a single file. Do note, though, that when you use this method, you must choose to export your dataset either as a CSV or TSV file (the PDF, XML, and CSV for Excel options will not respect this particular method to increase the limit value).

Table 1
Revised Finding Aid Statistics: Test Sample July 2007 - Oct 2010
(1-6 Webpage "Unique Page Views" or Less During 2008-2009)

COLL #	Cu. Ft.	FY 2007-08	FY 2008-09	FY 2009-10	July 2009 *	Aug. 2009	Sept. 2009	Oct. 2009 *	Nov. 2009	Dec. 2009 *	Jan. 2010 *	Feb. 2010 *	Mar. 2010	Apr. 2010	May 2010	June 2010	July 2010	Aug. 2010	Sept. 2010	Oct. 2010	TOTALS 7/2009 - 10/2010
#112	0.073	5	4	27	8	1	3	2	1	2	2	1	1	4	0	2	3	2	1	5	38
#162	0.073	9	6	26	5	1	5	1	4	5	0	2	2	1	0	0	1	1	1	1	30
#465	0.075	6	4	21	6	0	3	0	1	0	2	3	2	1	1	2	1	1	2	4	29
#517	0.22	9	6	49	3	1	5	7	11	1	12	2	2	3	2	0	1	1	3	1	55
#518	0.25	18	5	27	3	0	1	2	4	0	3	4	6	3	0	1	4	5	1	1	38
#409	0.35	16	1	118	0	0	0	1	0	0	0	0	0	0	99	18	7	7	3	3	138
#445	1.158	16	1	11	0	0	1	0	0	0	2	1	0	0	3	4	2	2	3	8	26
#579	0.43	12	1	10	0	1	0	0	1	0	0	2	1	2	2	1	1	1	2	1	15
#732	0.073	12	1	17	1	0	1	4	0	2	1	3	0	1	1	3	1	1	0	3	22
MG0063	0.001	17	1	8	0	0	0	0	0	1	0	1	0	0	2	4	2	3	0	1	14
TOTALS	2.703	120	30	314	26	4	19	17	22	11	22	19	14	15	110	35	23	24	16	28	405

Collections Tested with Web Addesses

#112	Exum L. Curl Papers http://digital.lib.ecu.edu/special/ead/findingaids/0112/
#162	George W. Lyon Papers http://digital.lib.ecu.edu/special/ead/findingaids/0162/
#465	Jeanie R. Little Papers http://digital.lib.ecu.edu/special/ead/findingaids/0465/
#517	Martha E. Donaldson Papers http://digital.lib.ecu.edu/special/ead/findingaids/0517/
#518	Democratic Women of North Carolina Collection http://digital.lib.ecu.edu/special/ead/findingaids/0518/
#409	Chatham Book Club Records http://digital.lib.ecu.edu/special/ead/findingaids/0409/
#445	District Bar Records Of The Second Judicial District Of North Carolina http://digital.lib.ecu.edu/special/ead/findingaids/0445/
#579	Clio Book Club Records http://digital.lib.ecu.edu/special/ead/findingaids/0579/
#732	Steward's Weekly Provisions Register http://digital.lib.ecu.edu/special/ead/findingaids/0732/
MG0063	John Vainwright Bible Records http://digital.lib.ecu.edu/special/ead/findingaids/MG0063/

* = Authors' UPVs during experiment.

COLL #	Cu. Ft.	FY 2008-09	FY 2009-10	July 2009 *	Aug. 2009	Sept. 2009	Oct. 2009	Nov. 2009	Dec. 2009	Jan. 2010	Feb. 2010	Mar. 2010	Apr. 2010	May 2010	June 2010	July 2010	Aug. 2010	Sept. 2010	Oct. 2010	TOTALS 07/2009 - 10/2010
#521	0.003	4	5	0	0	2	0	1	1	0	0	0	1	0	0	0	0	1	0	6
#572	0.073	4	9	0	0	3	0	0	1	0	0	0	2	1	2	0	0	1	0	10
#591	0.055	4	11	2	0	1	3	0	1	0	0	0	2	0	2	1	0	1	0	13
#606	0.403	4	5	0	0	0	2	0	0	0	0	0	1	0	2	17	1	1	1	25
#624	0.22	4	6	0	0	0	2	1	0	0	0	0	2	0	1	0	0	1	1	8
#677.031	0.203	2	2	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	3
#704	0.055	2	10	1	1	2	2	1	0	0	1	1	1	0	0	0	1	0	1	12
#820	0.23	2	3	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	2	5
#864	0.48	2	5	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	6
#987	0.001	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2
TOTALS	1.723	30	57	4	2	8	9	4	3	1	2	3	12	1	8	18	2	5	8	90

Control Sample & Web Addresses

#521	Eugene M. Ransom, Jr. Papers http://digital.lib.ecu.edu/special/ead/findingaids/0521/
#572	William J. Henning Papers http://digital.lib.ecu.edu/special/ead/findingaids/0572/
#591	Francesco Costagliola Papers http://digital.lib.ecu.edu/special/ead/findingaids/0591/
#606	Edward Crawford Williams Album http://digital.lib.ecu.edu/special/ead/findingaids/0606/
#624	Zach D. Cox Papers http://digital.lib.ecu.edu/special/ead/findingaids/0624/
#677.031	U. S. Navy Memorial Foundation Collection \ Joseph Vetter Papers http://digital.lib.ecu.edu/special/ead/findingaids/0677-031/
#704	F. Warren VanWert Papers http://digital.lib.ecu.edu/special/ead/findingaids/0704/
#820	Henry J. Conger Papers http://digital.lib.ecu.edu/special/ead/findingaids/0820/
#864	Branch & Company, Bankers Collection http://digital.lib.ecu.edu/special/ead/findingaids/0864/
#987	St. Thomas Episcopal Church Collection http://digital.lib.ecu.edu/special/ead/findingaids/0987/

* = Author's UPVs

Table 3
Changes to Online Finding Aids 2009-2010[1]

COLL. #	COLLECTION NAME	BIOG / HIST NOTE	PRELIM INVENT[3]	SCOPE NOTE	ACCESSION INFO	TOTAL WORDS ADDED
112	Exum L. Carl Papers*		1337			1337
162	George W. Lyon Papers*	201				201
465	Jeanie R. Little Papers	144	344			488
517	Martha E. Donaldson Papers[2]	220	512	40		772
518	Democratic Women of North Carolina Collection		110		47	157
409	Chatham Book Club Records	1266	764	140		2170
445	District Bar of the Second District of North Carolina Records		769		58	827
579	Clio Book Club Records		2163		54	2217
732	Steward's Weekly Provisions Register		87			87
MG63	John Vainwright Bible Records		127	88	35	250
		457.75	690.33	89.33	48.5	

Average of 851 words added per collection.

[1] Numbers indicate words added to online finding aids. These values are calculated by removing the EAD tags and then using the "text statistic" function within NoteTab Pro (version 6.12) to calculate a "word count."

[2] Three catalog subject entries added were also added at this time, which are not represented in the total word count figure.

[3] In the case of the "preliminary inventory" sections, not only were the XML tags removed, but the "container" and "unitid" tag values were also removed (as such, element values such as "1.a" – indicating Box 1, folder a – were not counted in the "word count").

[*] Indicates that this collection has also been cataloged with Library of Congress Subject Headings. However, the number of additional "words" contained within these subjects is not represented in the total word count figure.

References

[1] Referred to below generically as "archival," "archivists," "repositories," or "staffs."

[2] The quality of much of this data is questionable, especially the manually counted data. This is because repositories generally use staff members of varying reliability and training to generate such data.

[3] A related question that the authors did not seek to answer is: Are researchers using some collections because the web site was more easily accessible or easier to navigate?

[4] North Carolina Exploring Cultural Heritage Online. Prof. Dembo served as principal investigator for the grant project 2001-2003.

[5] The grant project also encoded guides for oral history, miscellaneous genealogy, map, and church records collections.

[6] At the end of FY 2008/2009, the EAD website produced 8,527 online pages which received a total of 83,387 UPVs. See Table 4.

[7] For more details on the development of the statistics cited in this paper, see Notes on Technical and Statistical Methods, by Mark Custer, below (Page 16).

[8] Follow this link for a list of the 114 least used online finding aids used in this experiment: http://ead2002.pbworks.com/Least+viewed+collections+online+%28for+2008-09%29. Professor Dembo selected the first and last five finding aids belonging to collections containing less than one cubic foot of material as a test sample; he selected the immediately adjacent five similar finding aids in each group as a control sample.

[9] See Table 1 Revised Finding Aids Statistics: Test Sample, July 2008-Oct. 2010. Prof. Custer later added statistics for FY 2007/08 (however, these statistics were culled from web server logs and not from Google Analytics, so they are not able to be used for any valid comparisons). The reason that collections #162, #517, and #518 are listed with more than 4 UPVs is because the original list of collections excluded a small subset of UPVs, which have subsequently been added. The balance of the online pages included staff directories, hours of operations, current exhibits, and descriptions of collection areas, policies and procedures statements, search engines, and links to other departments, among others.

[10] Subsequent investigation revealed that some of these collections received 1 or 2 more UPVs than first believed. The authors do not believe that this materially affects the conclusions of the research. See Table 1.

COLL #	COLLECTION NAME & ONLINE ADDRESS	WEBPAGE RANKING 2008-2009	WEBPAGE RANKING 2009-2010	RANKING CHANGE	WORDS ADDED TO FINDING AIDS
#409	Chatham Book Club Records http://digital.lib.ecu.edu/special/ead/findingaids/0409/	3,046	476	2,570	2,170
#732	Steward's Weekly Provisions Register http://digital.lib.ecu.edu/special/ead/findingaids/0732/	3,061	817	2,244	87
#579	Clio Book Club Records http://digital.lib.ecu.edu/special/ead/findingaids/0579/	3,055	1,120	1,935	2,217
#MG0063	John Vainwright Bible Records http://digital.lib.ecu.edu/special/ead/findingaids/MG0063/	3,072	1,673	1,399	250
#517	Martha E. Donaldson Papers http://digital.lib.ecu.edu/special/ead/findingaids/0517/	1,732	409	1,323	772
#445	District Bar Records Of The Second Judicial District Of North Carolina http://digital.lib.ecu.edu/special/ead/findingaids/0445/	3,047	1,797	1,250	827
#162	George W. Lyon Papers http://digital.lib.ecu.edu/special/ead/findingaids/0162/	1,728	644	1,084	201
#112	Exum L. Curl Papers. http://digital.lib.ecu.edu/special/ead/findingaids/0112/	1,727	731	996	1,337
#465	Jeanie R. Little Papers http://digital.lib.ecu.edu/special/ead/findingaids/0465/	1,731	918	813	488
#518	Democratic Women of North Carolina Collection http://digital.lib.ecu.edu/special/ead/findingaids/0518/	1,484	712	772	157
	AVERAGES	2,368.3	929.7	1,438.6	850.6
	TOTAL NO. OF ONLINE PAGES	8,527.0	12,492.0		
	TOTAL NO. OF UNIQUE PAGE VIEWS	83,387.0	87,084.0		

[11] Table 1: Revised Finding Aids Statistics: Test Sample, July 2007 – Oct. 2010 and Table 2: Revised Finding Aids Statistics: Control Sample, July 2008 – Oct. 2010 also show the cubic feet, names and web addresses of finding aids involved in the experiment. The tables also show the number of UPVs each finding aid received during the experiment. Readers interested in learning what the tested finding aids looked like revision should examine the finding aids in the control sample.

[12] Table 1 and Table 2 record the number of cubic feet in each collection and the overall volume of the collections included in the experiment.

[13] The UPV data highlighted in gold reflect the authors' access of the finding aids during the project.

[14] See Table 3: Changes to Online Finding Aids, 2009-2010. This table does not include the Control Sample finding aids since they remained unchanged during the experiment.

[15] The control sample received a total of only 90 UPVs for the entire 16 months of the test period which was higher than the 30 UPVs per year they received in FY 2008/2009. It was far less than the 25.31 UPVs per month received by the tested sample. The control sample received an average of 5.63 UPVs per month which was far less than the average received by the tested sample: 25.31 UPVs per month.

[16] The FY 2009/2010 statistics are inflated by the authors' own access of the website (highlighted in gold on Table 1) and by internally generated UPVs in May 2010 for Collection #409. Eliminating these statics would reduce the yearly numbers of UPVs to 176 and the yearly average to 17.6 UPVs per finding aid per year.

[17] Collection #606 in July 2010. That collection received 20 of the total 33 UPVs earned by all 10 finding aids in the control sample.

[18] Table 2 also shows the names and web addresses of each of the control finding aids. Readers interested in comparing the tested finding aids ranked before and after the experiment invited to visit these finding aids.

[19] Table 4: Ranking Changes to Tested Finding Aids, July 2007 – June 2010, shows that the Special Collections website had 8,527 pages in 2008/2009. This increased to 12,493 in 2009/2010, an increase of 31.74%. Meanwhile, the Special Collections website received a total of 83,387 UPVs in 2008/2009. This increased to 87,084 UPVs in 2009/2010, an increase of only 4.25%.

[20] Table 3: Changes to Online Finding Aids, 2009/2010.

[21] None of the finding aids fell in rank during the experiment.

Table 5
Impact of Changes to Finding Aid Elements, 2009-2010

COLLECTION NO.	WORDS ADDED TO FINDING AID ELEMENTS				Rank Change 2009-2010	Rank Change Per Word Added	Average Rank Change Per Word
	Biographical / Historical Note	Scope Note	Accession Information	Inventories			
#162	201				1,084	5.39	
#465	144				813	5.65	
#517	220				1,323	6.01	
#409	1,266				2,570	6.28	5.84
#112				1,337	996	0.74	
#465				344	813	2.36	
#517				512	1,323	2.58	
#518				110	772	7.02	
#409				764	2,570	3.36	
#445				769	1,250	1.63	
#579				2,163	1,935	0.89	
#732				87	2,244	25.79	
#MG063				127	1,399	11.02	6.16
#517		40			1,323	33.08	
#409		140			2,570	18.36	
#MG063		88			1,399	15.89	22.44
#518			47		772	16.43	
#445			58		1,250	21.55	
#579			54		1,935	35.83	
#MG063			35		1,399	39.97	28.45
TOTALS	1,831	268	194	6,213			

[22] Table 4: Ranking Changes to Tested Finding Aids, July 2009 – June 2010.

[23] Table 4: Ranking Changes to Tested Finding Aids, July 2009 – June 2010.

[24] Table 4: Ranking Changes to Tested Finding Aids, July 2009 – June 2010.

[25] Table 5: Impact of Changes to Finding Aid Elements, 2009-2010 shows the number of words added to each of four finding aid elements, including Biographical / Historical Note, Scope Note, Inventories, and Accession Information. It tracks the subsequent rank change each finding aid experiences and also records the rank change per word added for each finding aid. The table then shows the average rank change per word for each group of finding aid elements.

[26] In the case of Joyner Library's Special Collections Department, the authors estimate that it takes approximately 15 hours per cubic foot of manuscript material to completely "process" the collection. By this definition, it would have taken one staff person 40.55 hours to process (accession, arrange, describe, conserve and catalog) these materials, or more than twice the time actually taken in during this experiment.

[27] UPVs do not appear on the main "dashboard" of the Google Analytics interface. Instead, they are reported in the "Top Content" report of Google Analytics.

[28] This variable of 30 minutes, however, can be adjusted. Our UPVs are governed by a timeout value of the default 30 minutes.

[29] For more information about Google Analytics and how it uses a combination of cookies to identify users and sessions, see the following website: http://code.google.com/apis/analytics/docs/concepts/gaConceptsCookies.html

[30] For instance, Google Analytics' unique visitor tracking cookie, "_utma", has an expiration date of 2 years. But someone could configure their browser so that those cookies are deleted whenever they close their browser.

[31] In fact, one of our collections received a significant, but temporary spike in usage when it was included as an example in an email that was sent out to a wide listserv audience.

[32] For more information on how to apply filters, including a predefined filter based on IP addresses, see the following: http://www.google.com/support/analytics/bin/topic.py?hl=en&topic=11091