**Annual Statistics – Fiscal Year End**

Each year, in early July we will refresh the training server with the full back up done at end of day, June 30th. We will then merge Lists to create a list to hold the items added that year. We will notify staff to NOT load records to the training server until we have completed running the item counts by location and format type. We will run all our statistics against the training server. This will allow us to get numbers for the correct full fiscal year and not impact the production server. (Note: current fiscal year refers to the current fiscal year we are producing statistics for, not the fiscal year we are presently in.)

Here are the steps for the various types of counts we can get from Millennium:

**Prep**

**Before running any statistics, make the “personal manuscript” change (Dave).**

After refresh of the Training server implement a new item type so that annual stats can distinguish between Theses/dissertations and (other) Personal Manuscripts as required by CDL. These changes will be made to Production server at a later date.

- Add new item type 30, "Personal Manuscript"

- Change name of item type 27 from "Personal Manuscript" to "Print thesis/dissertation"

- Run a create list and gather all items with iType 27.

-Change items with biblevel not = "t" from itype 27 to itype 30.

- Retain itype 27 in items with biblevel = "t".

**UCB Record Work**

**Print Records**

**Item counts by Location and Format Type (Jackie):**

There is an additional sheet that has instructions for running the crosstab queries in Millennium, http://digitalassets.lib.berkeley.edu/annual\_statistics/documentation/Running\_crosstab\_queries\_for\_annual\_statistics.docx.

1. Run an Item Record crosstab report in Millennium Statistics against the full database by location and type (IType). Export the results as “Sheet1.xls”. Delete item type 23.
2. Run a create list of items suppressed or withdrawn (icode2 not equal to -) and not item type 23 (online resources) in the entire database. Run a crosstab report in Millennium Statistics by location and type (IType) against this new list. Export the results as “Sheet2.xls.”
3. Run a create list gathering item records with an item created date of the current fiscal year (e.g., July 1, 2015 – June 30, 2016) and not item type 23 (online resources). Run a crosstab report in Millennium Statistics against this set of records by location and type (IType). Export the results as “Sheet3.xls.”
4. Run a create list gathering item records that are suppressed or withdrawn (icode2 not equal to -) with an item created date of the current fiscal year and not item type 23 (online resources). Run a crosstab report in Millennium Statistics against this set of records by location and type (IType). Export the results as “Sheet4.xls.”
5. Run script to subtract the # 2 counts from the #1 counts. This will be the number we use total items. [allmillstats.pl]
6. Run script to subtract the #4 counts from #3 counts. This will be the number we use for items added this fiscal year. [fymillstats.pl]
7. Run script to subtract #6 counts from #5 counts to give us two final crosstab reports to process as described below: [before\_fymillstats.pl]
   1. Item totals for everything before the current fiscal year
   2. Item totals for everything added the current fiscal year

The output for each of these will be a crosstab report by Millennium locations and types which is not a one-to-one match with what has been requested (see <http://www.lib.berkeley.edu/Staff/CS/xls/ARL_UCOP_formulas_Under_Millennium.xls>)

For types, counts will be rolled up as shown in Appendix A: Type Mapping. Display the type in the left hand column with the counts.

For locations, counts will be rolled up based on the first 2 letters of the location code for UCB locations and the first 4 letters for NRLF locations, that is all locations starting with n. Display the 2 or 4 letter code with the counts.

**Withdrawn counts (Dave)**

A number of create lists will be run to determine the withdrawn counts. See Appendix B: Withdrawn Create Lists. Millennium Statistics will be run against these lists to create crosstab by location and type.

Withdrawn in lieu of, and contributed to DSP will also get crosstab reports.

Blank label locs; output the INT NOTE only.

If possible, output will be processed by script written to roll up counts (see above).

**Circulation Statistics (Dave)**

Checkouts, Renewals, and Reserves counts will be created by running Millennium Web Management Reports. Patrons will be as what shows in Appendix D: Patron Types.

Additional crosstab reports requiring post processing. Create two Excel sheets reporting circulations by patron type and location, one excluding reserves (circ\_rollup\_onlyres.pl), the other only reserves (circ\_rollup\_nores.pl). Millennium cannot mix record types in the statistics module. So in order to get patron type/location as the axis for the crosstab, a web management report needs to be generated using user specified date, patron type and location (See appendix E). The two scripts require location codes, web reports only provides the location “meaning.” The location codes are added by importing the report into an Access database (annual\_circ.accdb) and running a query joining a location table with the report table on the “meaning” (location description). The table then needs to be de-duped. That table is output as an Excel spreadsheet and the two scripts are run against it for the final crosstab reports. (**Note**: the “Loc\_Code” and “Location” field need to have their names changed to “Code” and “Meaning” for the scripts.)

**In House Statistics (?Dave?)**

Gather raw stats from Web Management Reports. Roll up the sub locations into the base locations using queries in the Access database located in O:\My Documents\ILS\Statistics\AnnualStats\_FY1\_FY2\InHouse directory.

**Total number of bib records, by location.**

Run statistics against the bib file and export a report based on location. Mount it on preservation assets.

**Electronic Records**

**Online Resources--E-Resources (Eileen)**

A number of create lists will be run to determine the online resources counts. See Appendix C: Online Resources Counts. These numbers are not grouped by location.

**Digital Collections statistics (Kurt)**

For digital collections we count the number of projects, the number of digital items and the amount (megabytes) of space those items use. We do not count METS files, but do count everything from image files (jpg, tiff) to audio to plain text and video.

A spreadsheet that lists all directories in /data/pa and /data2/da (on ucblibrary4) that we want to count and the library that they are counted for needs to be updated with the directories added in the last year. This spreadsheet is at <pathname>.

Once the spreadsheet is updated, a WPG programmer can run the script(s) that does all the heavy lifting of counting the files, etc. The script is located at <pathname> on ucblibrary4.

**Patron Records**

**Percentage of UCB patrons no-circ (Dave)**

Gather two lists, one of all current UCB patrons. Run the second list against the first gathering all records for patrons that have never checked out a book.

1st list-- all patrons with Patron Code1 = “0” (UC Berkeley) and an expiration date greater than or equal to today.

2nd list-- run against first list, and gather all the records for those patrons that had Total Checkouts = 0. Calculate percentage between two.

Put result in: https://preservationassets.berkeley.edu/annual\_statistics/percent\_circ/

**NRLF Record Work**

**NRLF cumulative deposit statistics (Dave)**

Gather all items from NRLF.

**Things we cannot count**

**Number of Pieces:** Millennium can count items, not how many physical pieces those items may represent. In Millennium, the item corresponds to the circulating unit: volume, box, carton, dvd, etc. The only way to count these in Millennium would be to add a field to the item record which staff would need to manually fill in with the correct number of pieces. We could then extract this field and add it all together to get the count. We do NOT recommend this as we think it will be a nightmare to maintain, not to mention back fill. Another option is to create a ratio based on the statistics provided in 2012. We could take the number from the Millennium reports for Microfilm, Microcards, Microfiche, and Microprints and compare it to the number of pieces reported for end of fiscal year 2012. Further investigation on this option is recommended.

**Transferred counts**: The “Next Steps on ARL-UCOP Annual Reporting – Counts & Circulations” is incorrect in that the new procedures for transfers is not yet in place, it has been proposed and is under review. Until then it will not be possible to use Millennium to gather these counts.

**On the fly item creation vs new items:** There is no way for Millennium to distinguish between new items and items that are created on the fly after they have been completed and cataloged, as both have a created date that will fall in the current fiscal year. Something would need to be added to the item record when it is created to distinguish between the new and created on the fly. This would need to be vetted by various Library groups.

**Serial titles counts:** Millennium Statistics doesn’t distinguish between print Serial Titles received currently purchased and not purchased. The above process will provide a count for serial items by location. We do not know how these were previously gathered, and would suggest consulting with Jim Gordon.

**Government Documents:** [When we did for statistics for last year it was determined to be too wide a count, so we will not be doing it this year] Gov Docs is not a format like the others on the list. Numbers for this will need to be gotten by a series of queries. See Appendix E: Gov Docs Create Lists. However, because these are not types like the others, they will be counting the same items counted by type. That is, a gov doc monograph will be counted as a monograph (volume) AND as a gov doc.

**And things we no longer do for annual statistics**

**NRLF Barcode Gaps**

/home/lsomgr/lib/BarcodeGaps/getAllBarcodes. Takes no arguments. Finds barcode gaps at NRLF and results can be viewed here. Uses Create lists slot 116 so that slot needs to be available prior to running (or edit script to use another available 1,000,000 record slot).

<http://ucblibrary3.berkeley.edu:8080/NrlfBarcodeGaps/>

**PublicPCs**

We are counting the number of times the public PC home page is served up for the Public PC. This runs against the Apache logs on cluster3 (access\_log.www<date>). Before running we need to verify the names of the home page script, e.g., bioshome.cgi. LCIS can supply an up-to-date list of Public PC IP addresses as a spreadsheet containing at least the columns “IPAddress”, “LocalMachineName”, and “LIb”. A README file with instructions on how to run the script is at ucblibrary4:/data/pa/annual\_statistics/publicPCs.

Appendix A: Type Mapping

|  |  |
| --- | --- |
| **Types to report** | **Which ones to count** |
| 1. Volumes | 1 Monograph - regular loan + 2 Monograph - short-term loan + 20 Atlas + 3 Bad Code + 0 Undefined + 47 Bad Code + 86 Reserves 1 Month + 87 Reserves 1 Week + 88 Reserves 1-day + 89 Reserves 1-day no renewal + 90 Reserves 2-hour + 91 Reserves 2-hour no renewal + 92 Reserves 2-hour no overnight + 93 Reserves 2-hr no renew/overnight + 96 Mellon Research Grant + 35 Score |
| 2a.Serials received currently - purchased/2b.Serials received currently - not purchased | 10 Serial-Periodical (regular loan) + 11 Serial - Periodical (other loan) + 15 Newspaper |
| 3a. Personal manuscripts | 27 Personal Manuscript |
| 3b. UC archival manuscripts | 28 UC Archival Manuscript |
| 3c. Other archival materials | 29 Other Archival Material |
| 4. Maps | 19 Map + 70 Aerial Photograph |
| 5a. Microfilm reels | 44 Microfilm |
| 5b. Microcards | 43 Microcard |
| 5c. Microfiche | 42 Microfiche |
| 5d. Microprints | 45 Microprint |
| 6. Pamphlets | 26 Pamphlet |
| 7. Government documents | n/a |
| 8a. Audiodiscs | 51 Phonograph Record/Audio disc LP |
| 8b. Audiocasettes | 53 Audio Cassette |
| 8c. Audioreels | 54 Audio Reel |
| 8d. Compact discs, digital audio | 52 Compact Disc + 55 Sound Recording, MP3, Undefined |
| 9a. Videotapes | 63 Videotape |
| 9b. Videodiscs | 62 Videodisc |
| 9c. Mult-media kits | 82 Multi-Media Kit |
| 9d. Motion pictures | 64 Film |
| 10a. Filmstrips | 65 Filmstrip |
| 10b. Pictorial items | 67 Visual Material: pictorial |
| 10c. 35mm slides | 66 Slide |
| 11a. Computer tapes | 75 Computer File |
| 11b. Monographic CD-ROM discs | 4 Monograph CD-ROM disc |
| 11c. Serial CD-ROM discs | 14 Serial CD-ROM disc |
| 11d. Floppy disks | n/a |
|  |  |
| **We will provide counts for the following even though they are not on the official list:** | |
| 23 Online Resources |  |
| 68 Visual Material: non-pictorial |  |
| 69 Visual Format Undefined |  |
| 76 Computer Hardware |  |
| 77 Media Equipment (audio/visual) |  |
| 81 Kit |  |

Appendix B: Withdrawn Create Lists

**[Change fiscal year as appropriate]**

CUMULATIVE WITHDRAWALS

Save item records

item status = "w", OR

item location starts with "zw"

Fiscal Year WITHDRAWALS TOTAL

Search CUMULATIVE file

Save item records

item internal note starts with "201507", OR

item internal note starts with "201508", OR

item internal note starts with "201509", OR

item internal note starts with "20151", OR

item internal note starts with "2016"

A crosstab spreadsheet using Label Location and iType will be generated for items withdrawn during the fiscal year being counted.

WITHDRAWALS (IN LIEU of STORAGE)

Search Fiscal Year TOTAL

Save item records

item internal note contains "in lieu of storage"

A crosstab spreadsheet using Label Location and iType will be generated for items that were withdrawn in lieu of storage.

WITHDRAWALS (CONTRIBUTED TO DSP)

Search Fiscal Year TOTAL

Save item records

item internal note contains "contributed to DSP"

WITHDRAWALS (EXISTS IN DSP)

Search Fiscal Year TOTAL

Save item records

item internal note contains "exists in DSP"

The Label Locations and iType for items counted in “contributed to DSP” and “exists in DSP” will be exported and included in an email to Gail (or that person in charge of annual statistics).

WITHDRAWALS (FED DOCS)

Search Fiscal Year TOTAL

Save item records

item internal note contains "Fed Doc"

A crosstab spreadsheet using Label Location and iType will be generated for items that were withdrawn for Fed Docs project.

REMAINING WITHDRAWALS

Fiscal Year TOTAL - (IN LIEU of STORAGE + CONTRIBUTED TO DSP + EXISTS IN DSP + FED DOCS) = REMAINING WITHDRAWALS

**Blank Location**

The Itype and Internal Note fields will be exported for all withdrawn items that are missing a Label Location.Appendix C: Online Resources Counts

Appendix C: Online Resources Counts

**CUMULATIVE e-RESOURCE SERIALS**

Save bib records as “CUMULATIVE E-SERIALS”

item location = "wb", AND

(bibliographic biblvl = "b", OR

bibliographic biblvl = "i", OR

bibliographic biblvl = "s"), AND

bibliographic MARC 85640 OR

bibliographic MARC 85641 OR

bibliographic MARC 9594

Dedupe file

**CUMULATIVE e-RESOURCE SERIALS PAID**

Use the “CUMULATIVE E-SERIALS” review file as the base

Save bib records as “CUMULATIVE E-SERIALS, Paid”

BIBLIOGRAPHIC  MARC Tag 85640  has  "Restricted to UCB IP"    OR BIBLIOGRAPHIC  MARC Tag 85641  has  "Restricted to UCB IP"    OR BIBLIOGRAPHIC  MARC Tag 9594  has  "Restricted to UCB IP"    OR BIBLIOGRAPHIC  MARC Tag 85640  has  "UCB Access"    OR BIBLIOGRAPHIC  MARC Tag 85641  has  "UCB Access"    OR BIBLIOGRAPHIC  MARC Tag 9594  has  "UCB Access"    OR BIBLIOGRAPHIC  MARC Tag 8569  has  "Restricted to UCB,"    OR BIBLIOGRAPHIC  MARC Tag 9599  has  "Restricted to UCB,"    OR BIBLIOGRAPHIC  MARC Tag 8569  has  "Restricted to UC C"    OR BIBLIOGRAPHIC  MARC Tag 9599  has  "Restricted to UC C"

dedup

**To determine the number to give them for “exclude from ARL stats from the above e-Resource Serials, paid (total)”**

Use the “CUMULATIVE E-SERIALS, Paid” review file as input

Save bib records

BIBLIOGRAPHIC  MARC Tag 799  has  "CRL Batch Record Load"

(this number is somewhat problematic since in some cases, there are TWO or more links in a given record -- CRL + SCP access or CRL + other paid access)

To eliminate the "combined" cases, using this file as input, search on:

BIBLIOGRAPHIC  MARC Tag 799  At Least one Field doesn't have  "CRL Batch Record Load"

Then subtract:  4,621 (paid; has CRL) - 1,091 (paid; has CRL, but also has other paid link) = **2,930**

**CUMULATIVE e-RESOURCE SERIALS FREE**

CUMULATIVE SERIAL - CUMULATIVE SERIAL PAID = CUMULATIVE SERIAL FREE

**CUMULATIVE e-Resource Serials, free  (total) = CUMULATIVE e-Resource Serials (total) minus CUMULATIVE e-Resource Serials, paid (total)**

(This number could very well be problematic, since some of these resources could be ones we pay for, yet were loaded long ago, before we started using the "Restricted to UCB IP" or "UCB Access" 856 labels. Yet this is the best we can do. Many do appear to be MARCIVE materials or other Gov Docs, so the assumption is that these are Freely Available as the default.)

**To determine the number to give them for “exclude from ARL stats from the above Cumulative e-Resource Serials, free (total)”,**

Use “CUMULATIVE E-SERIALS” review file as base

Save bib records

BIBLIOGRAPHIC  MARC Tag 599  has  "Hathi" OR

BIBLIOGRAPHIC  MARC Tag 856  has  "Marcive" OR

BIBLIOGRAPHIC  MARC Tag 793  has  "Open access resource; selected by the UC Libraries"OR

BIBLIOGRAPHIC  MARC Tag 799 has "US Office of Scientific and Technical Information"

=============================================================================

=============================================================================

**FY e-Resource Serials (total)**

Use the “CUMULATIVE E-SERIALS” review file as the base

Save item records as “FY\_SERIALS\_Items”

item created date between <dates of fiscal year, e.g., "07-01-2016" and "06-30-2017">

gather bibs attached to the saved file of item records “FY\_SERIALS\_Bibs”

**FY e-Resource Serials (total), paid**

Use the ““FY\_SERIALS\_Bibs” review file as the base

Save bib records

bibliographic MARC 85640 OR

bibliographic MARC 85641 OR

bibliographic MARC 9594

has "UCB IP" OR

has “UCB ACCESS”

dedup

**To determine the number to give them for “exclude from ARL stats from the above FY e-Resource Serials, paid”**

Use the “FY\_SERIALS\_Bibs” review file as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 799  has  "CRL Batch Record Load"

(this number is somewhat problematic since in some cases, there are TWO or more links in a given record -- CRL + SCP access or CRL + other paid access)

To eliminate the "combined" cases, using this file as input, search on:

BIBLIOGRAPHIC  MARC Tag 799  At Least one Field doesn't have  "CRL Batch Record Load"

Then subtract:  4,621 (paid; has CRL) - 1,091 (paid; has CRL, but also has other paid link) = **2,930**

**FY e-Resource Serials (total), free**

FY e-RESOURCE SERIALS FREE

FY SERIAL - FY SERIAL PAID = FY SERIAL FREE

**To determine the number to give them for “exclude from ARL stats from the above FY e-Resource Serials, free (total)”,**

Use the “FY\_SERIALS\_Bibs” review file as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 599  has  "Hathi" OR

BIBLIOGRAPHIC  MARC Tag 856  has  "Marcive" OR

BIBLIOGRAPHIC  MARC Tag 793  has  "Open access resource; selected by the UC Libraries"OR

BIBLIOGRAPHIC  MARC Tag 799 has "US Office of Scientific and Technical Information

=============================================================================

=============================================================================

**CUMULATIVE e-RESOURCE MONOGRAPHS**

Save bib records as “CUMULATIVE E-MONOGRAPHS”

item location = "wb", AND

(bibliographic biblvl = "a", OR

bibliographic biblvl = "c", OR

bibliographic biblvl = "d", OR

bibliographic biblvl = "m", OR

bibliographic biblvl = "t"), AND

bibliographic MARC 85640 OR

bibliographic MARC 85641 OR

Dedupe file

**CUMULATIVE e-RESOURCE MONOGRAPHS PAID**

Use the “CUMULATIVE E-MONOGRAPHS” review file as the base

Save bib records as “CUMULATIVE E-MONOGRAPHS, Paid”

bibliographic MARC 85640 OR

bibliographic MARC 85641 OR

has "UCB IP" OR

has “UCB ACCESS”

dedup

**To determine the number to give them for “exclude from ARL stats from the above CUMULATIVE e-Resource Monographs, PAID (total)”**

Use the ‘CUMULATIVE E-MONOGRAPHS, Paid” review file as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 799  has  "CRL Batch Record Load"

(this number is somewhat problematic since in some cases, there are TWO or more links in a given record -- CRL + SCP access or CRL + other paid access)

To eliminate the "combined" cases, using this file as input, search on:

BIBLIOGRAPHIC  MARC Tag 799  At Least one Field doesn't have  "CRL Batch Record Load"

Then subtract:  4,621 (paid; has CRL) - 1,091 (paid; has CRL, but also has other paid link) = 2,930

**CUMULATIVE e-RESOURCE MONOGRAPHS FREE**

CUMULATIVE MONOGRAPH - CUMULATIVE MONOGRAPH PAID = TOTAL MONOGRAPH FREE

**To determine the number to give them for “exclude from ARL stats from the above CUMULATIVE e-Resource Monographs, Free (total)”**

Use the “CUMULATIVE E-MONOGRAPHS” review file as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 599  has  "Hathi" OR

BIBLIOGRAPHIC  MARC Tag 856  has  "Marcive” OR

BIBLIOGRAPHIC  MARC Tag 793  has  "Open access resource; selected by the UC Libraries"  OR

BIBLIOGRAPHIC  MARC Tag 799 has "US Office of Scientific and Technical Information"

**CUMULATIVE e-RESOURCE TOTAL**

CUMULATIVE SERIALS + CUMULATIVE MONOGRAPHS = CUMULATIVE TOTAL

**FY e-RESOURCE MONOGRAPHS**

Use the “CUMULATIVE E-MONOGRAPHS” review file as the base

Save item records as FY\_MONO\_Items

item created date between <dates of fiscal year, e.g., "07-01-2016" and "06-30-2017">

gather bibs attached to the saved file of item records as FY\_MONO\_Bibs

**FY e-RESOURCE MONOGRAPHS PAID**

Use the “FY\_MONO\_Bibs” review file as the base

Save bib records as “FY\_MONO\_Bibs\_Paid”

bibliographic MARC 85640 OR

bibliographic MARC 85641 OR

has "UCB IP" OR

has “UCB ACCESS”

dedup

**To determine the number to give them for “exclude from ARL stats from the above FY e-Resource Monographs, PAID (total)”**

Use the “FY\_MONO\_Bibs\_Paid” review file as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 799  has  "CRL Batch Record Load"

(this number is somewhat problematic since in some cases, there are TWO or more links in a given record -- CRL + SCP access or CRL + other paid access)

To eliminate the "combined" cases, using this file as input, search on:

BIBLIOGRAPHIC  MARC Tag 799  At Least one Field doesn't have  "CRL Batch Record Load"

Then subtract:  4,621 (paid; has CRL) - 1,091 (paid; has CRL, but also has other paid link) = 2,930

**FY e-RESOURCE MONOGRAPHS FREE**

FY MONOGRAPH - FY MONOGRAPH PAID = FY MONOGRAPH FREE

**To determine the number to give them for “exclude from ARL stats from the above FY e-Resource Monographs, Free (total)”**

Use the “FY\_MONO\_Bibs” review files as the base

Save bib records

BIBLIOGRAPHIC  MARC Tag 599  has  "Hathi" OR

BIBLIOGRAPHIC  MARC Tag 856  has  "Marcive” OR

BIBLIOGRAPHIC  MARC Tag 793  has  "Open access resource; selected by the UC Libraries"  OR

BIBLIOGRAPHIC  MARC Tag 799 has "US Office of Scientific and Technical Information"

**FY e-RESOURCE TOTAL**

FY SERIALS + FY MONOGRAPHS = FY TOTAL

Appendix D: Patron Types

Alumni

Botanical Garden Docent

CA Private Univ/Coll Student

CA Public School Teachers

CA Resident

CA University/Colleges Faculty

CA University/Colleges Student

Dept of Health Services (DHS)

Executive/Management

Faculty/Academic Spouse

Faculty/Academic Staff

Far West Labs (WestEd)

GTU Staff/Visiting Scholar

Graduate Student

ICSI/MSRI/Howard Hughes Inst

ILL Other Libraries

ILL RLCP Institutional

ILL UC

LBNL/LLNL Academic Staff

LBNL/LLNL Non-Academic/Post-doc

Library Departmental Card

Library Development

Library Staff

NRLF Departmental Card

NRLF Google Cart

Non-Academic Staff

Non-Academic Staff Spouse

Other UC Spouse

Reference Use Only

Stanford Faculty

Stanford Graduate Student

UC Extension Faculty

UC Extension Student

UC Office of the President

UC Visiting Scholar/Post-doc

UCB Visiting Scholar/Post-doc

Undergraduate

Undergraduate - UCB SLE

Appendix E:

