Random Sample Generation

I. General

A. Purpose: Generate random samples for each collection (location) and create "pull lists" as well as review files for each collection based on sublocation codes. The Systems Department will generate samples based on "primary" sublocation codes and Preservation will generate samples based on "secondary" sublocation codes.[1]

B. The total size of each random sample will be determined using a sample size calculator and will be based on a 95% confidence level and a confidence interval of 3. Estimated sample sizes range from 667 to 1,066.

II. **Systems Department**: Random Sample Generation for "primary" sublocations.

A. The Libraries' Systems Department will create random samples for each primary sublocation code (e.g. norst) using the shelf list function of the Circulation\ Inventory module of Millennium (or Chinook) for most collections. The Circulation Department will provide assistance in developing random samples for smaller collections and for portions of collections housed at PASCAL.

B. Preservation will provide Systems with the information they need to generate this random sample.

1. Since the random sample function produces each sample list from a percentage of the entire shelf list, Preservation will convert the number required into a percentage, after allowing for the total number of items in each location (including all sublocations).

2. For instance, if Juvenile includes a total of 14,000 items and 2000 of those are oversize (juvov), the distribution of the random sample will be:

Juvst = 12,000
Juvov = 2,000
Total = 14,000

Random sample needed: 1,000 items. 1,000 ÷ 14,000 = .071 (7%)

C. Preservation will supply Systems with the call number ranges to be

[1] "Primary" sublocation code = the five character sublocation code which is assigned to the largest number of items within each location. For instance, juvst (but not juvov).
"Secondary" sublocation code = the five character sublocation code which is assigned to all items except those assigned the "primary" sublocation code within each location. For instance, juvov (but not juvst).

used for each location so that they can create a shelf list. Each list will be supplied 2-3 weeks before the random sample list is needed.

1. Shelf lists will include the entire range of call numbers for each location. Larger locations (such as Norlin Stacks) will be divided into smaller call number ranges and will require more than one random sample list.

2. Each shelf list will be limited to one five-character sublocation code (e.g. norst).

3. Once the shelf list has been produced, Systems will create a random sample and export that sample to an Excel file. The Excel file will include the following information:

Call Number	Title	Barcode (last 6 digits)
B128.C8 F73 2002	Confucius : the Gold	309610
B305.D44 A54	Diogenes: the story	374568

4. Preservation will convert the random sample list into a "pull list" that will be used for gathering items to be assessed. (See Collecting and Managing Samples from Collections.)

5. Larger collections will require additional steps since Millennium cannot produce a shelf list of less that one percent.

III. **Preservation Department**: Random Sample Generation for "secondary sublocations.

A. Millennium:

1. Search criteria:

a) Call number range that includes the first and last call numbers in the location scheduled for assessment.

b) Sublocation code(s) (juvov, norov, etc.) for the location. Sublocations with low holdings can be combined provided the call number range is applicable to all sublocations. Exclude "primary" sublocation code.

2. Consolidating sublocations within a location:

a) If separate review files were created for sublocations within a location, combine them into a single review file.

Append
Item
Record
Review
File
Search (if no criteria is specified, Millennium will append all items in the review file being searched.)

3. Sort the review file by call number.

4. Export the following fields from the review file:

Barcode
Location
Code
Call number
Volume #
Title (abbreviated – 245|a)

B. Creating Random Sample:

1. Open Excel file

a) Insert “sort digits” (fill column to left of barcode. These numbers will be used for the random sample and to sort the final list by call number)

b) Save the file as an .xls file.

2. Open file “TEMPLATE -RANDOM SAMPLES.xls”

a) Go to the worksheet labeled “loc code date”

b) Copy data range from the first file into the “ENTIRE” data range in the template.

ENTIRE -ORIGINAL DATA RANGE					
SORT DIGIT	Location code	BARCODE	CALL #(BIBLIO)	VOLUME #	TITLE
Curser here to paste data					

c) Rename the worksheet (tab) to show the sublocation

code and the date.

d) Save the template under a new name (e.g. “sublocations for {ART, SCI, etc.} .xls”)

3. Create named ranges (if existing ranges are inadequate):

SORTDIGITA
RANDOMA

4. Create random sample

Tools
Data Analysis
Sampling
“Input Range” = SORTDIGITA
Enter number of samples needed (based on percentage of entire location)
“Output Range” = RANDOM (range name – set for 400 rows) OK

5. Produce the Random Sample list.

a) Go to the “Random Sample List” worksheet.

b) Verify that the first item in the list is correct. This is produced by using VLOOKUP FUNCTION

Sort digit =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,1)

Loc code =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,2)

Barcode =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,3)

Call Number =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,4)

Volume # =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,5)

Title =VLOOKUP('LOC CODE DATE'!$A4,'TEMPLATE -RANDOM SAMPLES.xls'!ENTIRE,6)

d) Highlight the first row and copy the formula down (enough

rows for the entire sample).

d) Sort the list by the sort digit column.

e) Delete duplicates.

f) Save the file.

6. Convert to Pull list format.