Use Internet Archive’s [*Advanced Search returning JSON, XML, and more*](http://archive.org/advancedsearch.php) to get all of our volumes. You want to download the following columns into csv: identifier; identifier ark; contributor; scancenter, details.

Search:

Contributor:Boston College Libraries (2443)

Or: Collection:Boston\_College\_Library (2444) \*\*\*

Download into CSV format – allow more than 50.

Identify the new identifiers using access compare to find unmatched.

Put the new identifiers in a text file and run get-ia-metadata.pl to get the xml metadata from the fInternet archive.

Put a <records> wrapper around the resulting file and remove all those extra xml declarations.

Run the xslt hathi-meta.xml to extract a ppe delimited file of the meatadata needed for HT work. <xsl:value-of select="call\_number"/><xsl:text>BCL01|</xsl:text>  
 <xsl:value-of select="identifier"/><xsl:text>|</xsl:text>  
 <xsl:value-of select="identifier-ark"/><xsl:text>|</xsl:text>  
 <xsl:value-of select="identifier-access"/><xsl:text>|</xsl:text>  
 <xsl:value-of select="volume"/><xsl:text>|</xsl:text>  
 <xsl:value-of select="title"/><xsl:text>&#10;</xsl:text>

ID, DASIFileName, status, hathiReportStatus, hathiReportTime, note, identifier, aleph, ark, link, volume, title