Online Supplement

In Figure 1 basic and applied biological science includes biology, the biomedical sciences, and agriculture; physical science includes physics, chemistry, mathematics, earth sciences and engineering; and the social sciences includes economics, political science, sociology, psychology, history, linguistics, education and business. A conservative estimate of distinct manuscript pages was obtained by merging all documents with a given title. Serial estimates were obtained by matching OCLC serials to journals listed in ThompsonReuters’ Science, Social Science and Arts and Humanities Citation Index by issn. For the 7,982 matching issns, we estimated the year for volume one based on the sequence of volume numbers in later years, censoring if below 1750, and estimated Epanechnikov kernel densities for journal start-years in nine broad subject areas. Finally, we multiplied these start-year probabilities by the entire OCLC serial holdings in each of the nine areas, and multiplied by the average annual number of pages in those areas for the final estimate. Online journal page estimates were obtained by matching OCLC holdings to serials listed in Fulltext Sources Online’s (FSOs) by issn. FSO lists which online publishers provide each journal or if the journal is available freely through a publisher’s website, and for how many back-issues. For the 13,871 issns in OCLC and FSO, we estimated Epanechnikov kernel densities for start and end-year of online availability for nine broad subject areas. For each year, we subtracted end-year from start-year probabilities, multiplied by the average number of pages in the associated ThompsonReuters issns, and multiplied this quantity by the entire OCLC holdings, by broad subject.

Figure 1 shows the distribution of book holdings across libraries, but split by the function of the library—whether primarily reading (e.g., neighborhood, prison, and regional network libraries) or research (e.g., archives and university, national, and professional school libraries). It is clear that the reading libraries have many more books in common than the research ones. The spike at the end of the research library distribution traces the cannon of common books held at most such libraries.
Figure S1. Distribution of volumes across research and reading libraries: Number of volumes plotted against the number of OCLC libraries in which each are held, split by manuscript and serials, subject, and language. All distributions feature a spiking tail suggesting a core collection of books that appears in nearly all libraries of that type.