As academic and research libraries undertake large-scale collection realignment projects in order to reduce their collections’ physical footprint, government documents collections are often seen as ideal candidates for shrinkage. This is especially true for withdrawal criteria such as currency, which, in addition to usage, many federal publications both current and historic are available electronically from no-fee and subscription sources. Furthermore, maintaining tangibles government documents collections requires a significant financial and labor investment.

Unfortunately, institutions’ efforts have largely been stifled by incomplete holdings and usage data for their tangible materials. When libraries transitioned to OPACs in the 1970s and 1980s, government documents were not universally incorporated into the systems. Although great strides have been made in subsequent years, many libraries continue to have significant portions of their collections uncataloged. A byproduct of this oversight is that it limits the quantity and quality of the usage data that can be gathered on these materials. Because cataloging these materials and acquiring sufficient data demands a significant resource investment, many libraries undertake collection stewardship projects without sufficient data.

The purpose of this study was to harvest, analyze, and identify exactly electronic usage data of government publications could be used to make collections decisions. Previous studies have explored similar themes; however, limitations on the depth and breadth of usage data hindered efforts to develop a full understanding of library users’ government information use.

The study was conducted at the University of Maryland Libraries (UMD) in a large research institution in College Park, Maryland. The UMD is the regional federal depository library for Maryland, Delaware, and the District of Columbia. In addition to its tangible materials, the University Libraries provide online access to government information through federal and vendor resources.

For this study, vendor database usage statistics were harvested from usage reports from 2007-2016. The reports accessed contained data harvested from COUNTER and non-COUNTER compliant sources and had been aggregated for internal use. Because of changes in metadata naming and data collection practices over time, a thorough review of the data was required to eliminate duplication and misleading data. For example, "PerQuest session data did not correspond to a specific database, but rather combined session data for all products. Duplicating this data was necessary to avoid an artificially high usage count. However, this approach to data gathering for statistical research purposes, would be too time-consuming and labor-intensive. Therefore, the study focused on the usage data harvested through the OPAC, web, and specific databases.

Another issue with the vendor data was duplication. Database usage reports included data points for search terms, not specific titles. The most likely explanation for the spike was that specific terms were being used to search and discover content.

Analysis of the PURL data was easier than anticipated and a more in-depth analysis should be undertaken in the future. Despite difficulties analyzing the data, undertaking this type of project at least every five years can prove beneficial in the future. Recently researchers in the medical fields used text-mining to analyze literature in their field to identify information gaps. The existing data tables can be analyzed using more advanced statistical analysis programs such as R, SPSS, and SAS to gain more insights into users’ subject matter needs.

Data from vendors and the GPO can help libraries gain insight into users’ information needs. • Despite flaws in the data sample, both sets of data indicated users’ reliance on content published by the Legislative and Executive branches, which is widely recognized in the academic community. The reliance on this type of publication is expected to continue into the future. Researchers in the medical fields used text-mining to analyze literature in their field to identify information gaps. The existing data tables can be analyzed using more advanced statistical analysis programs such as R, SPSS, and SAS to gain more insights into users’ subject matter needs.

• **Use of vendor data will require more experimentation and exploration.** Due to inter-service variations in searching strategies and database features, it is difficult to develop a clear picture of database usage across different services. Despite difficulties analyzing the data, undertaking this type of project at least every five years can prove beneficial in the future. Researchers in the medical fields used text-mining to analyze literature in their field to identify information gaps. The existing data tables can be analyzed using more advanced statistical analysis programs such as R, SPSS, and SAS to gain more insights into users’ subject matter needs.

**Conclusion**

Data from vendors and the GPO can help libraries gain insight into users’ information needs. Despite flaws in the data sample, both sets of data indicated users’ reliance on content published by the Legislative and Executive branches, which is widely recognized in the academic community. The reliance on this type of publication is expected to continue into the future. Researchers in the medical fields used text-mining to analyze literature in their field to identify information gaps. The existing data tables can be analyzed using more advanced statistical analysis programs such as R, SPSS, and SAS to gain more insights into users’ subject matter needs.

Use of vendor data will require more experimentation and exploration. Due to inter-service variations in searching strategies and database features, it is difficult to develop a clear picture of database usage across different services. Despite difficulties analyzing the data, undertaking this type of project at least every five years can prove beneficial in the future. Researchers in the medical fields used text-mining to analyze literature in their field to identify information gaps. The existing data tables can be analyzed using more advanced statistical analysis programs such as R, SPSS, and SAS to gain more insights into users’ subject matter needs.