Forecasting Database Usage During the Height of COVID-19

Chris Sharpe and David Evans

Abstract:

As with other catastrophic events, the COVID-19 pandemic disrupted higher education and library services. The authors examine the effect of the COVID-19 pandemic on usage of health databases in 2020. They use time series analysis to create a forecast based on previous years' activities, and then compare it with actual database usage during the pandemic. The results show an initial increase in searches for the first full month of the pandemic, but then match the expected forecast data and decrease during summer and early fall months. The authors conclude that time series analysis is a useful tool for understanding the impact of events and for planning purposes.

Forecasting Database Usage During the Height of COVID-19

Chris Sharpe and David Evans
Kennesaw State University, USA

ABSTRACT

As with other catastrophic events, the COVID-19 pandemic disrupted higher education and library services. The authors examine the effect of the COVID-19 pandemic on usage of health databases in 2020. They use time series analysis to create a forecast based on previous years' activities, and then compare it with actual database usage during the pandemic. The results show an initial increase in searches for the first full month of the pandemic, but then match the expected forecast data and decrease during summer and early fall months. The authors conclude that time series analysis is a useful tool for understanding the impact of events and for planning purposes.

Keywords: COVID-19, Database Usage, Time Series Analysis, Health Science

INTRODUCTION

The COVID-19 pandemic caused great upheaval in lives and institutions, but catastrophic events are not new in the United States. Over 74 years ago, Gaus (1947) wrote about how catastrophic events have altered public policy. In the 19th century, outbreaks of malaria and yellow fever were common disruptive events in the coastal southern states. The 20th century saw two major global wars, a flu pandemic (1918), a global stock market crash (1929), and major events around HIV. More recently, events such as environmental disasters in the Philippines (1991 Mount Pinatubo), earthquakes in Sichuan, China (2008), the earthquake, tsunami, and nuclear accident in Fukashima, Japan in 2011, and local wars in 2019 in Kashmir, all serve as examples of local catastrophic events that affect higher education and library usage.

Since the beginning of the pandemic, much has been written both in the popular press and in academic literature about the disruption of COVID-19, which ranges from global supply chain problems to the delivery of education at the local level. Higher Education has responded with shifting to online-only learning or hybrid courses with minimal in-person seating. Library responses to the pandemic were immediate and, in many situations, initiated continuity plans that had been long in development. Some completely closed their buildings, while others remained open with restrictive access. Research help services had to switch to completely online, making use of reference chat services and video conference software such as Zoom or Microsoft Teams.

The purpose of this research was to examine the effect of COVID-19 on usage of select health-related databases. We did this by examining the patterns of usage in the immediate transition period (3-4 months) and the long-term period (6+ months) using time series analysis. A secondary purpose of our investigation was to attempt to confirm if pre-pandemic continuity

emergency plans for continuation of library operations worked. To our knowledge, no studies have explored how a traumatic event (e.g. 9/11; COVID-19) affects monthly database usage. To achieve this purpose, we explored two working questions: 1) Will a major national health event result in increases in database usage? 2) Will disruptions to normal campus activity result in a decreased usage of databases?

LITERATURE REVIEW

Jeong and Kim (2010) have produced an excellent annotated bibliography on time series analysis for librarians. Tenopir and Read (2000) reviewed database usage data from public and academic libraries and surveyed librarians on their views of database habits. They found a correlation between the number of workstations in the libraries and increased usage. Coombs (2005) tested the methodology of analyzing sessions of database usage using proxy server information to gather quantitative data and determine where and how the resources were being accessed. The results determined that there was a small percentage of databases being used, a preference for some full-text databases over others, and that students were primarily accessing the databases by name rather than by subject. A survey of faculty in a health science university in Tanzania regarding usage of e-resources and information literacy skills revealed a lack of awareness of what databases were available (Lwoga & Sukums, 2018).

Time series analyses can be used to predict periodic behaviors around events. McGrath (1996) used spectral analysis to understand cycles, periods, and frequencies in library collection usage. Normally collected by librarians in periods of semesters, years, months, or days, the analysis provided strong evidence for a seven-day week and semester activities and weak evidence of monthly cycles, while it also revealed an unexpected semiannual cycle. Murgai and Ahmadi (2007) used multiple regression to predict reference desk interactions to help forecast the number of staff needed at the desk. They used building traffic and semester time period as the predictor variables to relate to the dependent variable of the activity at the reference desk. Ahmadi et al. (2008) used an exponential smoothing model to further forecast library traffic and students coming to the reference desk.

How do major incidents such as natural disasters, pandemics, or war affect library use and research? Featherstone et al. (2012) examined how the influenza virus H1N1 pandemic of 2009 affected information needs among health care administrators, and how health sciences librarians could be successful in assessing and choosing material in support of administrative decision making. In a comparison of usage of electronic resources at a university during a period of conflict to a time of peace in Kashmir, there was a dramatic decrease in usage of e-resources and reduction of research output during the conflict years. Curfews, interruptions to attendance, and reduced internet connectivity affected access and research activities (Gul et al., 2014).

A survey of health science college teachers in Pakistan identified issues with teaching during the COVID-19 pandemic (Aziz et al., 2020). They included poor internet access, lack of experience with online teaching, concerns about the lack of hands-on learning, and confusion and difficulties with assessment. In Singapore, the availability of clinicians was a challenge, but the technology used to assist learning and communication resulted in a greater number of questions from students than normally experienced in-person (Cleland et al., 2020).

Health science researchers are greatly affected by lockdowns and other restrictions due to the pandemic. Conferences have been cancelled, in-person data collection is severely limited, clinical trials are placed on hold, and budgets are strained (Emans et al., 2020; Singh et al., 2020).

Many researchers rely on secondary data analysis of available datasets to continue scholarship efforts (Spurlock Jr., 2020). A series of focus groups of medical students in the Kingdom of Saudi Arabia found an overall positive experience in the switch to online synchronous learning, with most students being able to manage their school and family time well and enjoying the advantages of being able to re-watch lectures. The positive experience, however, depended on the type of classes (Khalil et al., 2020).

BACKGROUND

Kennesaw State University is a Carnegie designated, R2 institution and has two campuses. The enrollment in 2020 was around 41,000 students. The library system has access to 200,000 e-resources and more than 300 database subscriptions. There is an online library portal to authoritative, subscription-only information that is not available through free search engines or internet directories. The portal has a reporting tool for gathering usage statistics that acts in concert with the library system's central repository of usage statistics. This allowed us to create usage reports on all databases used, on specific databases or groups of databases, and on specific data elements for various periods of time. Table 1 illustrates the broad count of all database searches by library users in a given year. The upward annual trend in searches is evident until 2020. The university switched to online-only classes in the middle of March that year.

Table 1. Annual Electronic Database Searches across all Titles

Year	Searches
2016	2,393,735
2017	2,186,772
2018	3,302,096
2019	4,947,826
2020	2,421,858

Understanding the complex blocks of time within a semester has value to librarians with scheduling, purchasing resources, and allocating staff. Likewise, understanding what effect traumatic events (e.g. 9/11; COVID-19) have on usage of resources is important to librarians in their efforts with assisting students with library-related course work and faculty with their teaching and research efforts.

Kennesaw State University's academic year spans from mid-August of one calendar year to late July of the following year. An academic year is distinct from a calendar year in that an academic year has distinct blocks of time that are important to students, e.g. beginning semester, drop-add deadlines, last day to withdraw, mid-terms, final exams, and end of semester. Academic years can also exhibit seasonality such as holidays, e.g. Labor Day, Columbus Day; religious holidays, such as Thanksgiving, Hanukkah, Christmas, Passover, Easter; national days, such as MLK Day; or even Spring Break.

We collected the number of monthly searches conducted by library users from January 2016 through the end of March 2020 on usage of the databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Health Source: Nursing/Academic Edition, Medline with full text @ EBSCO, and Nursing & Allied Heath Database (ProQuest). These numbers were collated into an Excel spreadsheet. The combined annual searches for the selected databases

searches are illustrated in Table2.

Table 2. Annual Searches in Select Health Science Databases

Year	Searches
2016	55,115
2017	43,633
2018	63,901
2019	62,005
2020	60,847

METHODS AND RESULTS

Time Series Analysis is used extensively in business and medical research to forecast future events. An example is the CDC's often reported 7-day moving average of COVID-19 cases. Time series has four general areas that are examined: trend; a cycle component above/below that trend; seasonality (a short-term pattern); and an irregular event (which is everything other than trend, cycle, or seasonal events). Using a time-series model provided a forecast of expected database usage based on past performance. When a major event occurs (e.g. poor weather, seasonal flu outbreak, COVID-19) the actual results of a time period can be compared to the forecast to infer the impact of that event. In our investigation, we used a simple centered moving average (CMA) of two months to examine both seasonality and random events that might have an effect on usage. The difference in actual searches and CMA points toward random events that influence the number of searches. Upon speculation, these events might include such events as one-time large student activities or fluctuations in different assignments and due dates for research papers.
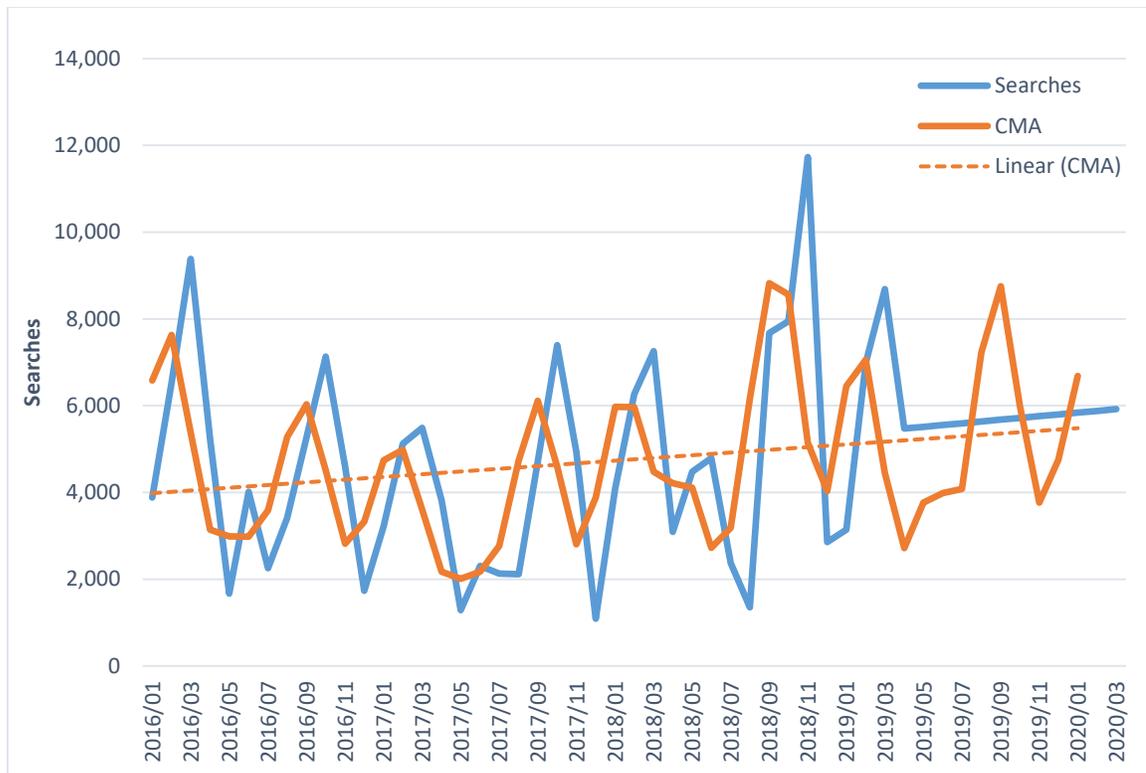
Figure 1. Seasonality and Search Trends of Health-Related Databases

The centered moving average was used to smooth the data in order to identify patterns by simplifying the random events between the time intervals. CMA can also be used in a limited fashion to forecast the next month usage. Figure 1 shows the seasonal movement of the searches. The highest number of searches occur in February, March, October, and November, or the second and third months of the spring and fall semester. The lowest number of searches occur during the summer semester months. The upward usage trend during this period is the dotted linear line. Part of our short-term analysis examined the immediate three months following the entire campus shutdown and courses moving to completely online. The 2020 search data followed the same declining usage from year to year at semester ends in all previous months of May. As the summer semester begins in June, the pattern trended upward as expected.
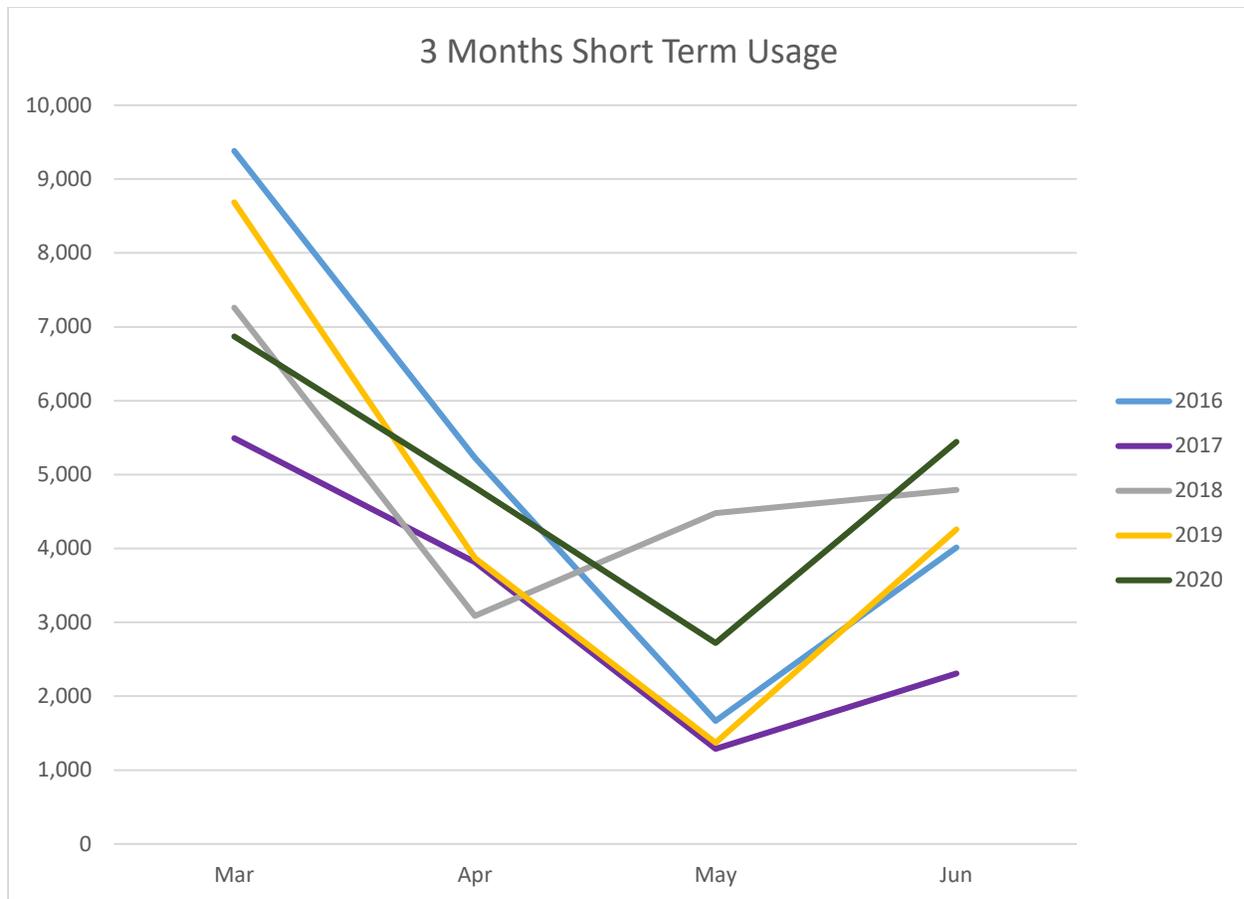
Figure 2. Actual Yearly Usage

For the most part, the forecasted searches where very close to the actual searches performed. Searches conducted in May, October, and November were dead on the mark. In April, the numbers were higher than the prediction. We are limited in discerning student "intent" around the use of these particular databases. Usage in these databases could be related to on-going pandemic research by nursing students or concerned library users' needs for information on COVID-19 symptoms and treatments. Usage might also be related to students working on assignments or just exploring the health literature related to pandemics in general.

The actual results were much lower than the forecast during the summer and the first couple of months of the fall semester. The campus returned to having in-person classes in Fall 2020, and so the decrease in searches may reflect the adjustment period during the pandemic. The searches are closer to the forecast and exceed it in the later months, which may indicate a return to pre-COVID-19 research activity.

Table 3. Long-term Forecast and Actual Searches

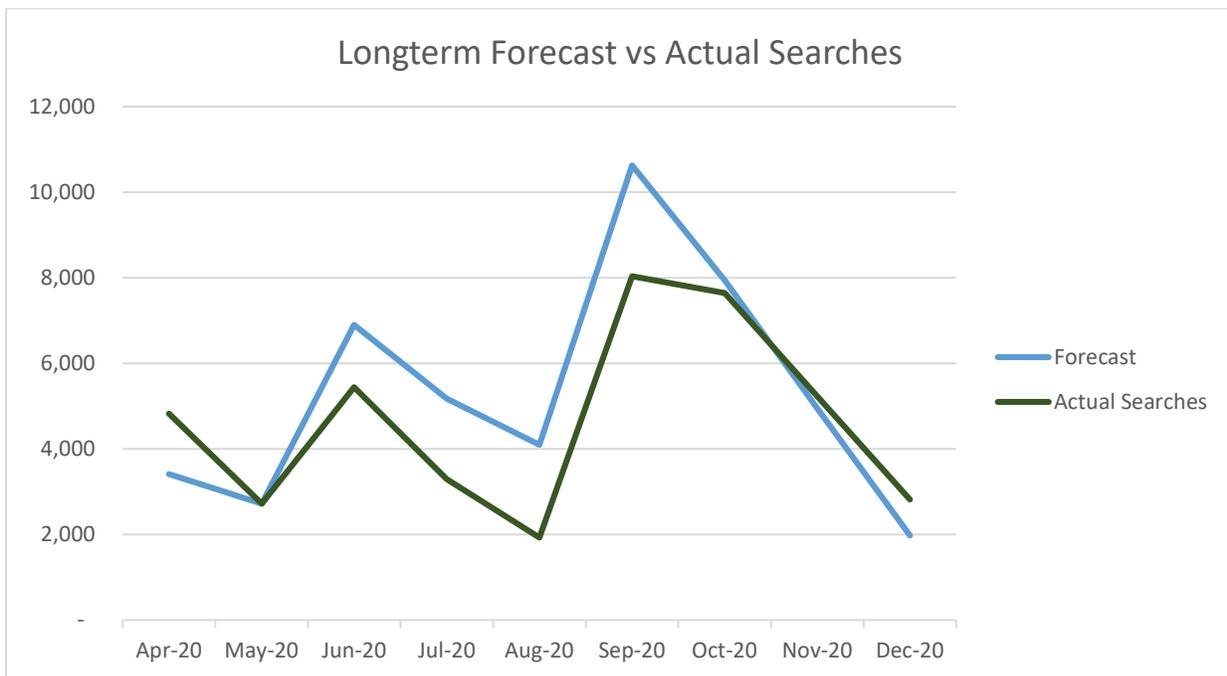| Month	Forecast	Actual
Apr-20	3,414	4,826
May-20	2,724	2,721
Jun-20	6,901	5,444
Jul-20	5,176	3,293
Aug-20	4,096	1,928
Sep-20	10,627	8,035
Oct-20	7,942	7,645
Nov-20	4,947	5,227
Dec-20	1,978	2,818

Figure 3. Long-term Forecast vs. Actual Searches

One interesting observation is that there was a decrease in library website traffic during the COVID-19 outbreak. We speculate that this observation was related to the course management

system directly connecting students to the databases and not going through the main library splash-page. Acknowledging this phenomenon is important when analyzing usage of the library, because a decrease in library website traffic may not necessarily equate to a decrease of library usage due to the various ways users can access library electronic collections.

CONCLUSION

Time series analysis can be a valuable method to understand and articulate the patterns and trends of library usage. What can we do with this forecast? One important lesson is that the forecast can help with decision-making regarding the allocation of resources for online research and staff support of those activities. Comparing the forecast with the actual results gives us an idea of the expected impact on databases if there is another significant event like the COVID-19 pandemic. Having this data will help us be better prepared for spikes or decreases in usage. An interesting study would be to see if an increase in database usage correlates with an increase in interlibrary loan requests. Also, a study on how well library continuity plans addressed the research needs of users during a catastrophic event would show best practices and challenges when creating and revising business continuity plans for higher education and libraries.

References

Ahmadi, M., Dileepan, P., Murgai, S, & Roth, W. (2008). An exponential smoothing model for predicting traffic in the library and at the reference desk. *The Bottom Line*, *21*(2) 37-48. https://doi.org/10.1108/08880450810898283

Aziz, A., Aamer, S., Khan, A. M., Sabqat, M., Sohail, M., & Majeed, F. (2020). A bumpy road to online teaching: Impact of COVID-19 on medical education. *Annals of King Edward Medical University, 26*, 181-186. https://annalskemu.org/journal/index.php/annals/article/view/3635

Cleland, J., Tan, E. C. P., Tham, K. Y., & Low-Beer, N. (2020). How Covid-19 opened up questions of sociomateriality in healthcare education. *Advances in Health Sciences Education, 25*(2), 479-482. http://dx.doi.org/10.1007/s10459-020-09968-9

Coombs, K. A. (2005). Lessons learned from analyzing library database usage data. *Library Hi Tech, 23*(4), 598-609. http://dx.doi.org/10.1108/07378830510636373

Emans, S. J., Ford, C. A., Irwin, C. E. J., Richardson, L. P., Sherer, S., Sieving, R. E., & Simpson, T. (2020). Early COVID-19 impact on adolescent health and medicine programs in the United States: LEAH program leadership reflections. *Journal of Adolescent Health, 67*(1), 11-15. https://doi.org/10.1016/j.jadohealth.2020.04.010

Featherstone, R. M., Boldt, R. G., Torabi, N., & Konrad, S. L. (2012). Provision of pandemic disease information by health sciences librarians: A multisite comparative case series. *Journal of the Medical Library Association, 100*(2), 104-112. https://dx.doi.org/10.3163%2F1536-5050.100.2.008

Gaus, J. M. (1947). Reflections on Public Administration. University of Alabama Press.

Gul, S., Ahmad Shah, T., & Ahmad, S. (2014). Digital user behaviour of academicians in a conflict zone, Kashmir: Comparing log analysis of electronic resources in the times of conflict and peace. *Program: Electronic Library and Information, 48*(2), 127-139. https://doi.org/10.1108/PROG-06-2013-0026

Jeong, S. H., & Kim, S. (2010) Core resources on time series analysis for academic libraries: A selected, annotated bibliography. *Proceedings of the Charleston Library Conference*, 229-238. http://dx.doi.org/10.5703/1288284314839

Khalil, R., Mansour, A. E., Fadda, W. A., Almisnid, K., Aldamegh, M., Al-Nafeesah, A., Alkhalifah, A., & Al-Wutayd, O. (2020). The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: A qualitative study exploring medical students' perspectives. *BMC Medical Education, 20*(1), 285. https://doi.org/10.1186/s12909-020-02208-z

Lwoga, E. T., & Sukums, F. (2018). Health sciences faculty usage behaviour of electronic resources and their information literacy practices. *Global Knowledge, Memory and Communication, 67*(1/2), 2-18. https://doi.org/10.1108/GKMC-06-2017-0054

McGrath, W. E. (1996). Periodicity in academic library circulation: A spectral analysis. *Journal of the American Society for Information Science, 47*(2), 136-145. https://doi.org/10.1002/(SICI)1097-4571(199602)47:2<136::AID-ASI5>3.0.CO;2-%23

Murgai, S. R., & Ahmadi, M. (2007). A multiple regression model for predicting reference desk staffing requirements. *The Bottom Line, 20*(2), 69-76. https://doi.org/10.1108/08880450710773002

Singh, J. A., Bandewar, S. V. S., & Bukusi, E. A. (2020). The impact of the COVID-19 pandemic response on other health research. *Bulletin of the World Health Organization, 98*(9), 625-631. http://doi.org/10.2471/BLT.20.257485

Spurlock Jr., D. R. (2020). Scholarship during a pandemic: Secondary data analysis. *Journal of Nursing Education, 59*(5), 245-247. https://doi.org/10.3928/01484834-20200422-02

Tenopir, C., & Read, E. J. (2000). Database use patterns in public libraries. *Reference & User Services Quarterly, 40*(1), 39-52. https://www.jstor.org/stable/20863899

About the authors

Chris Sharpe is the Director of Access Services at the Kennesaw State University Library System.

Dr. David Evans is the Dean of the Kennesaw State University Library System.