

VIUS REPORTS 8.7

Authentication Logs

From August, 2002 to the present, traffic on 3 images databases licensed by the University Libraries was monitored with authentication logging. A web page was interposed and users who had selected one of the databases were asked to provide a user ID before connecting. These IDs were then searched against a library patron file to identify status (faculty, student, staff, etc.) and departmental or campus affiliation. The IDs were then encrypted to protect confidentiality and so that future authentications by the same ID could be identified as repeat uses. (VIUS project staff IDs were removed from the data.) The 3 databases monitored were:

- The AMICO Library™ – a cooperative collection of more than 65,000 digital images of works of art from more than 30 museums, mostly in North America.
- The AP Multimedia Archive™ – a database of approximately 700,000 historical and current photos collected by the Associated Press.
- The VIUS/CONTENTdm™ Database – a database of approximately 12,000 images. This database came online August 29, 2002.

Authentication logging cannot provide enough context to answer the general question “Are these databases frequently used?” However, we can obtain some useful perspective by comparing traffic on these databases to the roughly 400 others offered by the University Libraries (primarily indexes and full-text resources). Both the AP Multimedia Archive™ and the AMICO Library™ were available at Penn State throughout the calendar years 2002 and 2003 and statistics were available for that period through September 2003. When the number of connections for those years for all University Libraries databases were ranked, these two pictorial databases stood well above the median (near the top of the second quartile) but well below the average (which was raised by half-a dozen databases with extremely high numbers). A respectable showing that might be described as “moderate.”

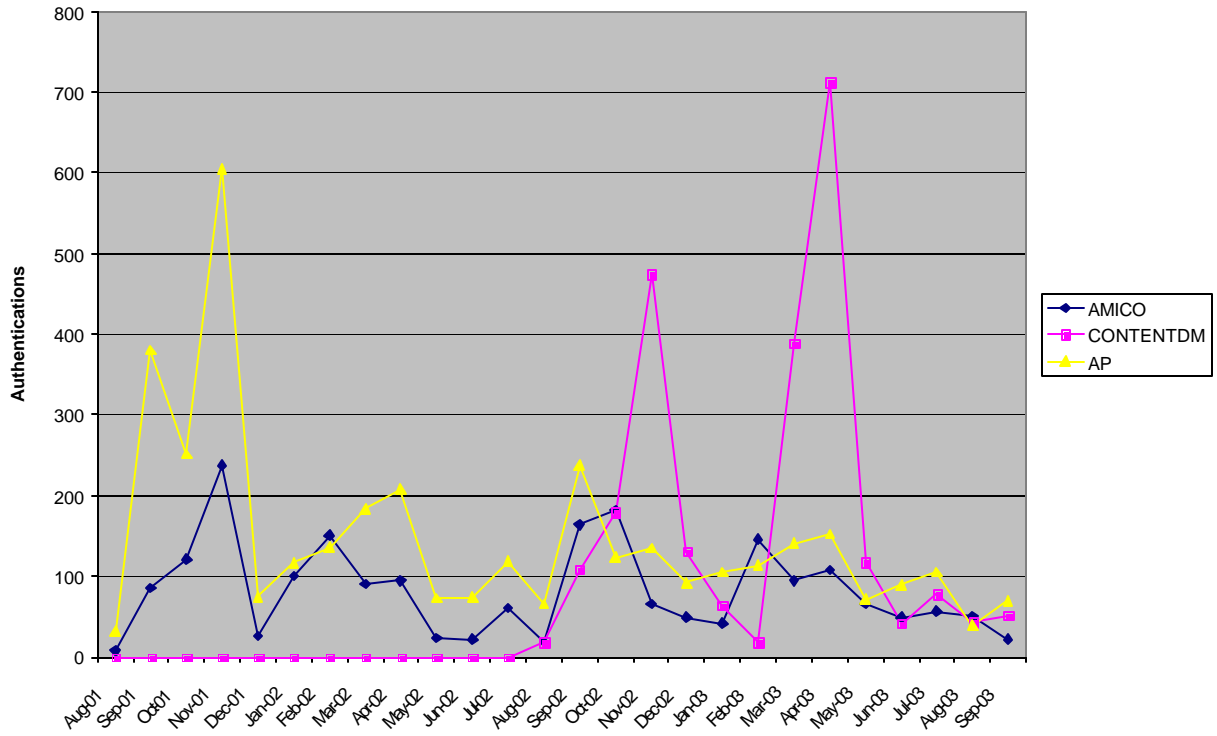
Turning to the authentication logs, we examined the data collected from August 29, 2001, through September 16, 2003. During this time 2,814 people authenticated on the picture databases 8,307 times (for an average of 2.95 authentications per user). Looking at the subset of this time when all three databases were available (August 15, 2002 – September 16, 2003), 1865 people authenticated 5022 times for an average of 2.69 authentications per user. This lower average is expected due to the shorter timeframe. We can then compare the authentication numbers for all three databases over the same period of time: AMICO 403 users/ 1112 authentications for an average of 2.76 per user; CONTENTdm™ 930 users/ 2387 authentications for an average of 2.57 per user; AP Multimedia Archive 657 users/ 1523 authentications for an average of 2.32 per user.

Also of interest is looking at the authentications per user ID over time. The 2 year averages are higher than the 1 year averages. However, they are not double but rather just somewhat higher. For instance, the overall authentications per user ID average went from 2.69 to 2.95 when doubling the amount of time. Additionally, by looking at the changes in the numbers of unique IDs over that time and the raw authentication numbers, it becomes clear that continued usage of the system over time was dominated more by new users who use it a couple of times rather than retaining previous users who will continue to use the system. In other words, there seems to be a revolving door of users who will use it two or three times and then be replaced by new users. This is true of both databases but is more true of AP where increasing the time by 100% increased the number of users by 111% versus AMICO where doubling the time increased users by 73%. Since overall usage hasn't increased dramatically (as seen in the following chart) it is possible to deduce that although many new users were added many must have been lost -- hence the revolving door theory. Of course, this is based on just over two years of data. It would be interesting to see this if this holds true over the longer run.

By looking at the total number of unique user IDs across all three databases compared to the subtotals within each database, it can be determined that crossover usage is fairly low. Out of 1865 users over the course of a year, an estimated 3.4% to 6.7% of users used more than one image database within that time frame.

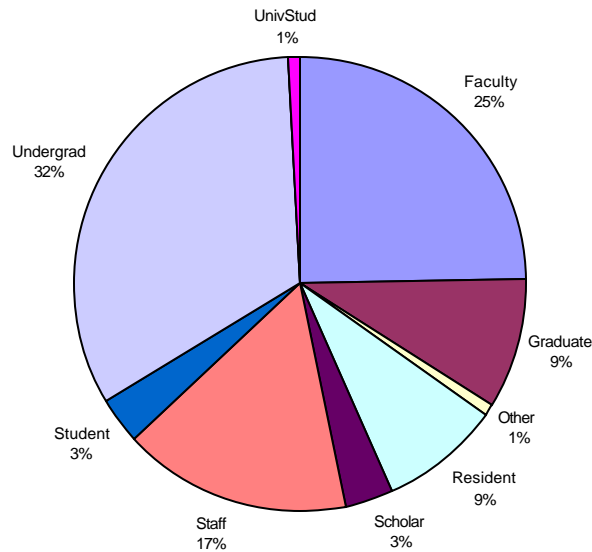
The following chart and table show significant fluctuations by month (apparently due to assignments or seasonal variations in class offerings, etc). The two peaks associated with CONTENTdm™ usage corresponds precisely with an extra credit assignment initiated in cooperation with the VIUS project. The assignment occurred in a fall and spring introductory Landscape Architecture class with enrollments of 700 students each semester. The assignment intended to promote usage of the CONTENTdm™ database in order to obtain informed opinions about the database for a survey taken in that class. It was completed by approximately 1/3rd of the students each semester. In addition, the most extensive publicity for the VIUS database (a large format color postcard) was distributed during the week of March 24, 2003. The fact that these stimulants show so clearly on the chart illustrates how sensitive database use is to matters such as assignments or publicity.

Database Authentications by Month (8/29/01 - 9/16/03)

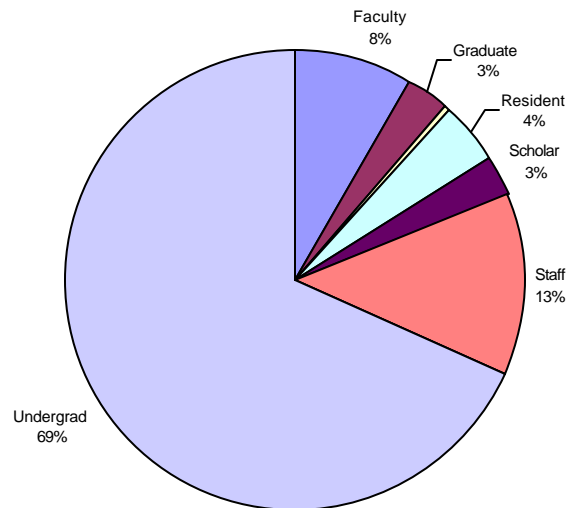


Considered by status, the authentication logs suggest that students constitute a larger percentage of authentications on the CONTENTdm™ database and AP Multimedia Archive™ than they do the AMICO Library™.

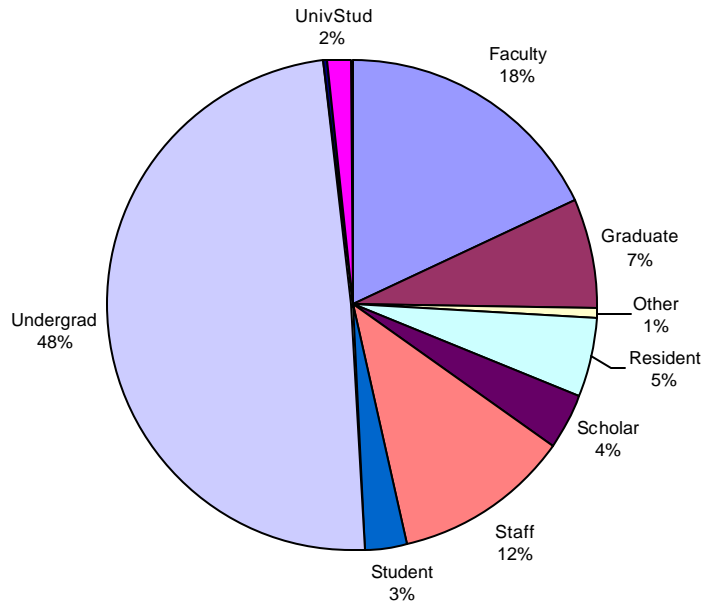
AMICO Authentications by Status (n=2113)
8/29/01-9/16/03



CONTENTdm Authentications by Status (n=2400)
8/15/02-9/16/03

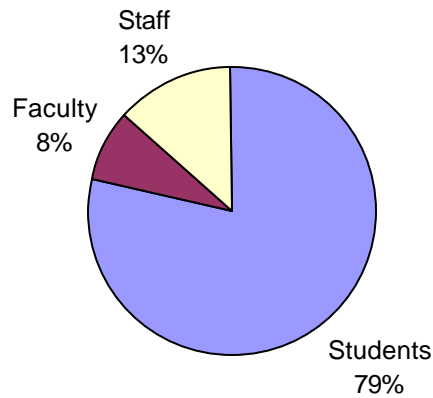


**AP Multimedia Archive Authentications by Status (n=3766)
8/29/01-9/16/03**



By comparing the demographics above to the generalized distribution of the Penn State population, we can see that faculty usage of the AP Multimedia Archive and in particular of the AMICO Library™, is higher than their mere size in the population would suggest.

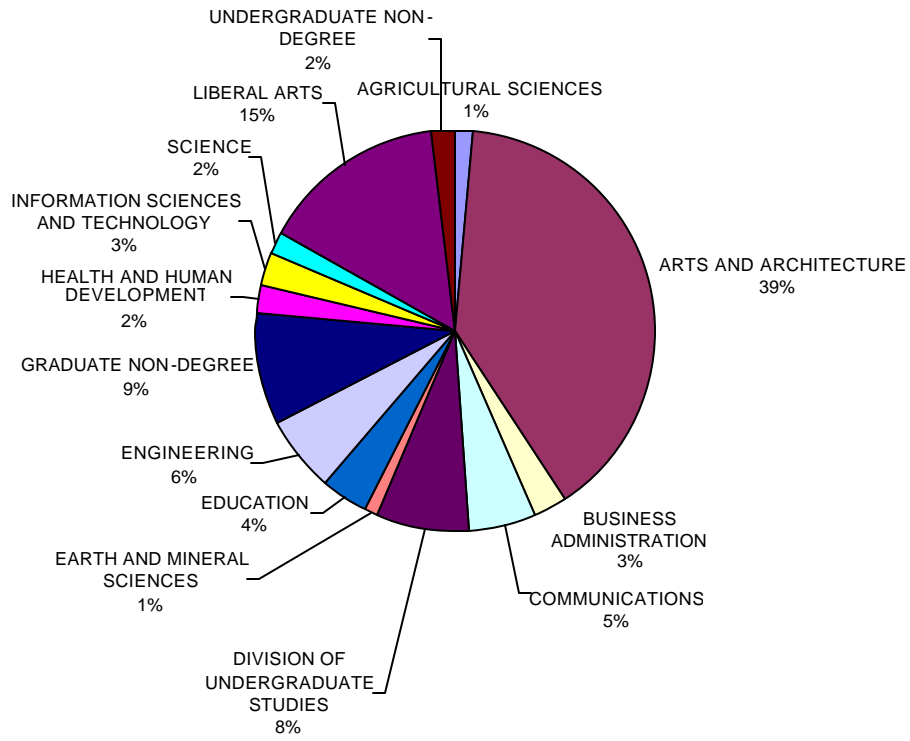
**Generalized Penn State Population
(n=approx. 95,500 people)**



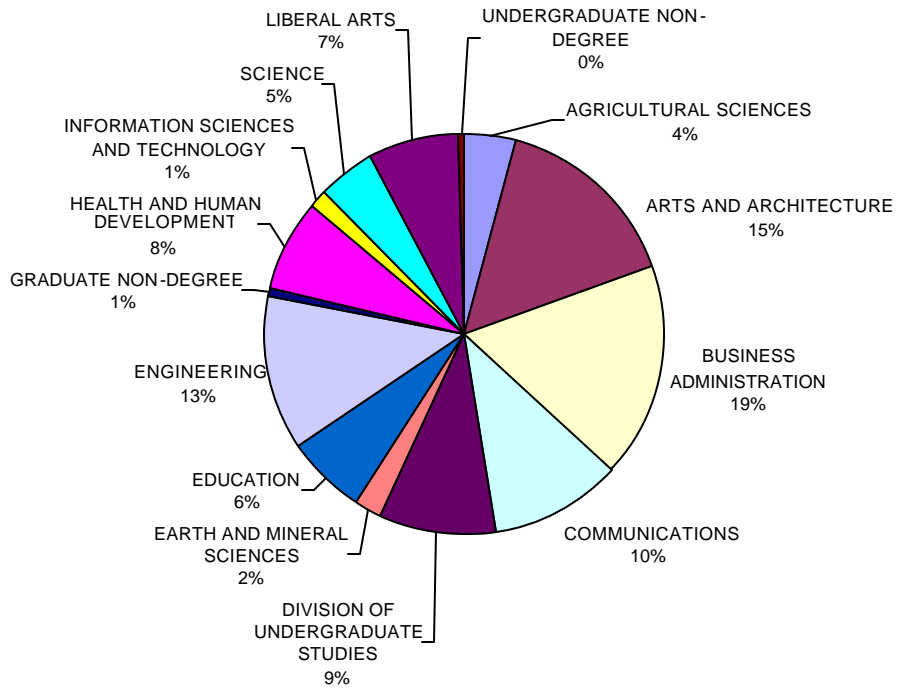
Analysis of the authentication logs by discipline is a confusing matter. Similar Penn State colleges have different names at some campuses. Moreover, many of the authentications could not be affiliated with any department because of address list problems. Nonetheless, by limiting authentications to those from the main campus, the following charts illustrate two aspects of use by discipline:

- The AMICO Library™ shows a larger percentage of uses by College of Arts and Architecture people than do the other two databases.
- The use of the VIUS/CONTENTdm database, even though its subject matter is specialized, probably illustrates the diverse disciplinary composition of the large Landscape Architecture course – a “general education” course which fills a requirement for most undergraduates.

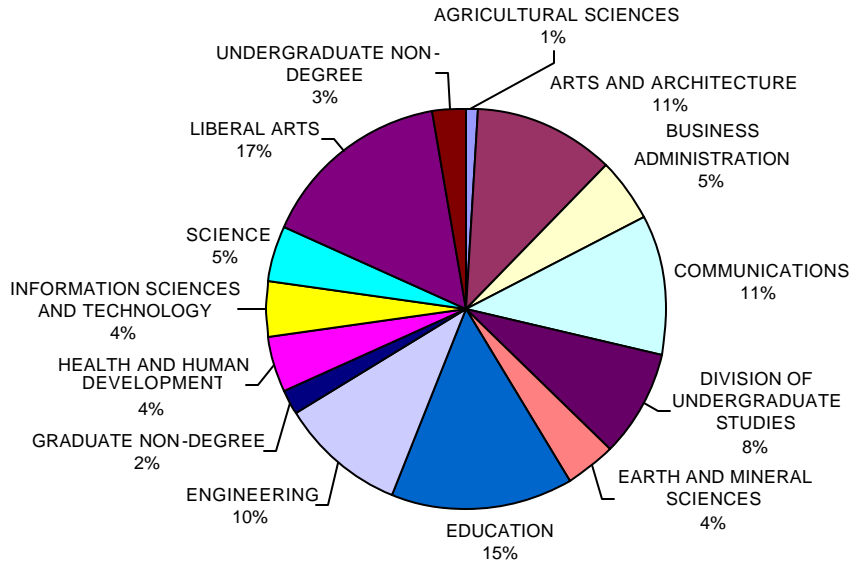
**AMICO Library™ Use with All Users by College at UP Campus
8/29/01-9/16/03 n=991**



CONTENTdm™ Use with All Users by College at UP Campus
8/29/01-9/16/03 n=1887



AP Multimedia Archive™ Use with All Users by College at UP Campus
8/29/01-9/16/03 n=2281



Jim Frost and Henry Pisciotta
October, 2003