Library Assessment Conference | Preconference Workshop
Data Analysis and Presentation, September 25 & 27, 2006

Workshop Survey—Summary Results

In late July, I surveyed 40 workshop registrants as an aid to my preparations. Twenty-seven people responded. Here are a few salient results.

ORGANIZATIONAL ROLES: Nearly all the respondents have a managerial or high level coordinating role with respect to data analysis or assessment within their libraries. (Is this an indication of a strategic shift in how research libraries operate?) One person has a supporting role in a department dedicated to measurement (a sign of significant commitment to this activity). Two reported committee assignments (I anticipated a larger selection of this category) and four had "other" roles with respect to data, such as web analytics in an IT unit. Two respondents play important roles but are **not** librarians, and of them one is a professional statistician employed as a consultant. (An interesting expansion of the professional competencies found in research libraries today.)

PRIORITIES: Here's how people responded to a question about the assessment and data collection priorities at their institutions (most work at public universities).

Assessment | Data Collection Priorities	Vote Cnt.	Rank	Pct. (n=27)
Service Quality	21	1	78%
Reference/Instruction	20	2	74%
Collection Use (print/digital)	20	2	74%
Organizational Accountability	17	3	63%
Strategic Planning	16	4	59%
Collection Management	14	5	52%
Organizational Efficiency/Productivity	12	6	44%
Finance and Personnel Administration	8	7	30%
Data Repository Development	8	7	30%
Other (describe)	5	8	19%

No. of respondents

No. of priorities selected
2
4
2
2
3
10
2
2
9
8
7
6
5
4
3
2

ISSUES PEOPLE ARE WRESTLING WITH:

Organizing and presenting data and making it useful for decision-making

Developing tools for analysis and data management and distribution

Collecting and making sense of vendor data

Getting the assessment effort off the ground

Dealing with cultural resistance

Systematizing data collection

Organizing and tracking projects

Integrating findings from multiple assessment initiatives

Coping with the time management demands of assessment–time constraints of data collection trump analysis

Cleaning the data: resolving inconsistency in the disparate lines of data reporting

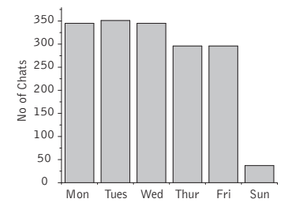
The chicken or the egg…"Others are looking at me to tell them what data are important for us to have, and at the same time I'm asking the question–'What is it you need to know for planning services and resources?'."

Just the overwhelming scope of it all

Notes on Graphical Presentation

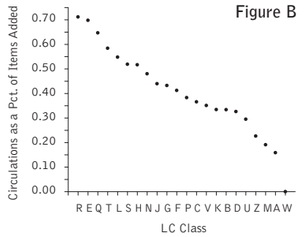
Graphs vs. Tables: Graphs help us visualize relationships, variation and trends, and patterns that may have explanatory force. Tables are best suited for small, non-comparative, and labeled data sets–which describes a lot of the data libraries handle. Figure A is a graph that should have been a table. It was used at Penn in a report about chat activity. In the same time it takes to decode the graphic, a table presents the same and even more information–information like frequencies and other descriptive statistics that took up additional graphs in the report. A table also makes it easier to spot potential problems in the data, like the multiple repetition of frequencies on Monday and Wednesday and on Thursday and Friday.

Figure A



Day	Count	Pct.	Cml. Pct.
Mon	345	20.7%	20.7%
Tue	351	21.0%	41.7%
Wed	345	20.7%	62.3%
Thr	296	17.7%	80.1%
Fri	296	17.7%	97.8%
Sun	37	2.2%	100%

Figure B



Sometimes the choice between a graph or a table is not clear cut. The graph in Figure B is a display of labeled data that started life as a table. The graphic seems to have an edge, in part because it reveals the regular distribution of the 22 data points. The graph plots circulation per items added by LC class in an 18 month period starting with January 2005.

Time Series: Time series are the essential tools for trend-spotting. Think about context and relationships when using time series plots and whether juxtaposition or superimposition is the most informative way to display your data. Figure C uses separate but quantitatively linkedplots to display change in Penn's e-journal and serials print holdings since 2000. These data could have been combined in one graphic, but as the magnitudes of the collections and the rates of change are very different, superimposition would sacrifice clarity and detail. In Figure D the opposite holds. The graph plots the use of Penn's electronic resources before and after a major redesign of the library web site. Here the superimposition of data lines is crucial to determining if visitors are having trouble navigating the site. The data plotted show logins just prior to and after the redesign, and compare these to logins at the same time in the previous year.

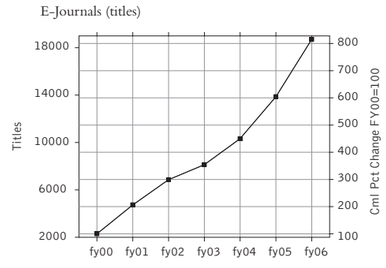


Figure D

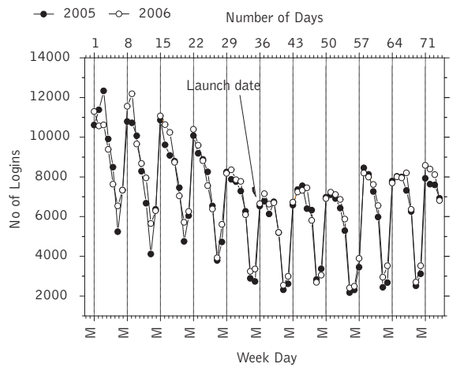
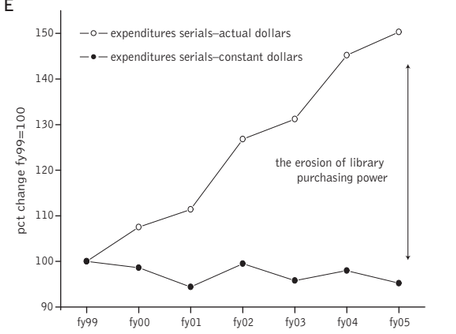
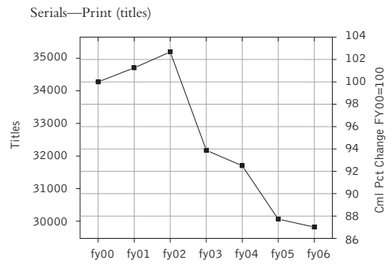


Figure E



Serial expenditures adjusted for inflation using the U.S. Periodical Prices Index, 1984-2005 (titles exclude Russian transalations)

Figure C



Graphs are Contextural: The time series is a device for measuring factor / response relationships where time is the factor. But in the time series examples given so far, temporal factors take a back seat to other aspects of the analysis. Think about how to put data most effectively into proper explanatory contexts when approaching graphical presentations. Figure E is an example of contextualizing the data. Here the graph plots the current dollar increase in serials spending at Penn, a time series that acquires new meaning when superimposed on a line that shows the same amounts adjusted for inflation.

Graphing Tips: Don't hesitate to include explanatory captions. As Edward Tufte put it, text is data too. Include sources: citations enhance the credability of data. Don't hesitate to start a scale at a point other than zero, (Figure C,) but be sure the scales are appropriately labeled so as not to confuse the decoding. Use multiple scales to enhance explanatory power of the graphic. In Figure C this happens by providing counts and percent increases and in Figure D by giving the x variable in days of the week and in a count of days. Spikes and troughs in time series can hamper the visual decoding of data; consider using moving averages to smooth out spikes, or logarithmic scales to normalize distributions that tend either to cluster around a measure or have values that vary by orders of magnitude. The visual analysis of data is an iterative process. Experiment with different graphing strategies and test them with people. At time two different approaches to a graph will reveal aspects of data that a single approach can miss.