Choosing the Best Tools for Evaluating Your Library

Abstract

Numerous library assessment tools are available today with proven validity and reliability. It is difficult to know if one has selected the *best* tool to build one's assessment case given one's unique circumstances. Answering the following series of questions will provide guidance. The questions are: 1) Why are we evaluating? 2) Who will use the results? 3) Do we have baseline or do we need to establish one? 4) What will this tool tell us and what is the precision of its measurement? 5) What new information will we gain? 6) What are the initial and continuing costs for using this tool? 7) What are the staffing requirements and what new operational knowledge does the staff gain? 8) Will the assessment resonate with and help support the goals of the library's parent organization? 9) How will the findings be used by the library's parent organization? 10) How might the findings from the assessment be used *against* the library? Tactics for answering these questions are provided, accompanied by graphic illustrations of the different paths one can take in the choosing the best library assessment tools for particular circumstances.

Introduction

Numerous library assessment tools are available today with proven validity and reliability. However, the answer to the librarian's question 'Which one(s) of these tools should I use to demonstrate the library adds value to our parent organization?' is not available. Answers to the following questions can provide an informed basis upon which to select the tool best suited for a given circumstance. *Figure 1. Decision Flowchart for Selecting Library Assessment Tools* is provided to help illustrate this network of questions and their relationships to each other.

A conceptual model, *Figure 2. Strategic Points for Library Performance Measurement* illustrates and provides examples of evaluation measures and the general points at which measures are taken. Tools for evaluating libraries are defined broadly and run the gamut from courts of inputs through evaluation of a program's impacts. They can be either quantitative or qualitative and measuring the number of new books added to a collection or the funds expended for those books in a year. It could be the way library customers feel about the service provided by library staff. The tool could calculate a rate of return on investment that derives from a public library or the impacts a university library's programs has on students about intellectual property rights. Integrated systems of evaluations such as the *balanced scorecard* which are being used by libraries are also included.

Today there are many evaluation tools being tested and utilized by libraries around the world. Examples of some of these are provided in *Figure 3. Library Assessment Environment, 2006*. Some of the organizations working in these areas are also noted in this figure which provides examples and is not comprehensive.

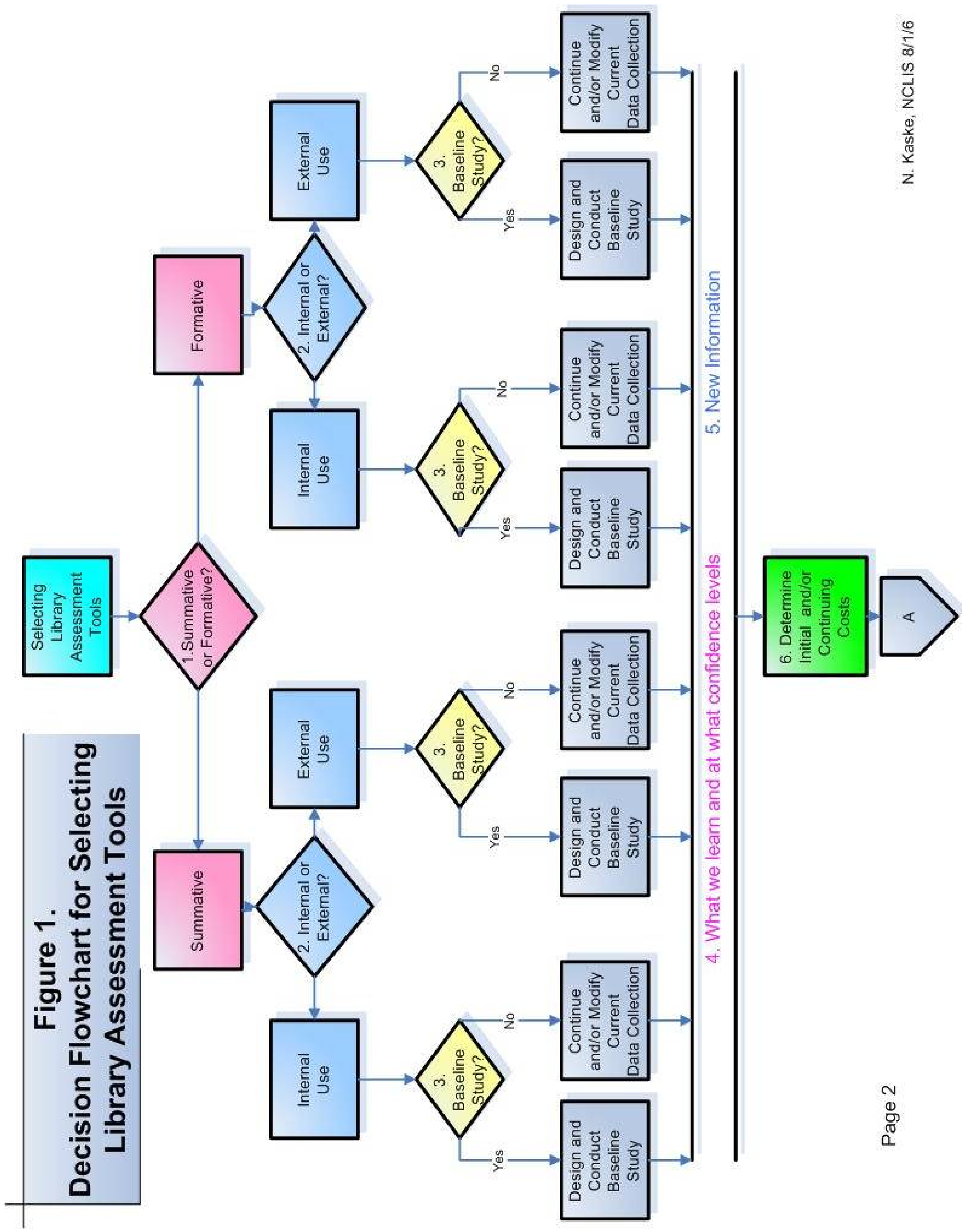

Figure 1.
Decision Flowchart for Selecting Library Assessment Tools

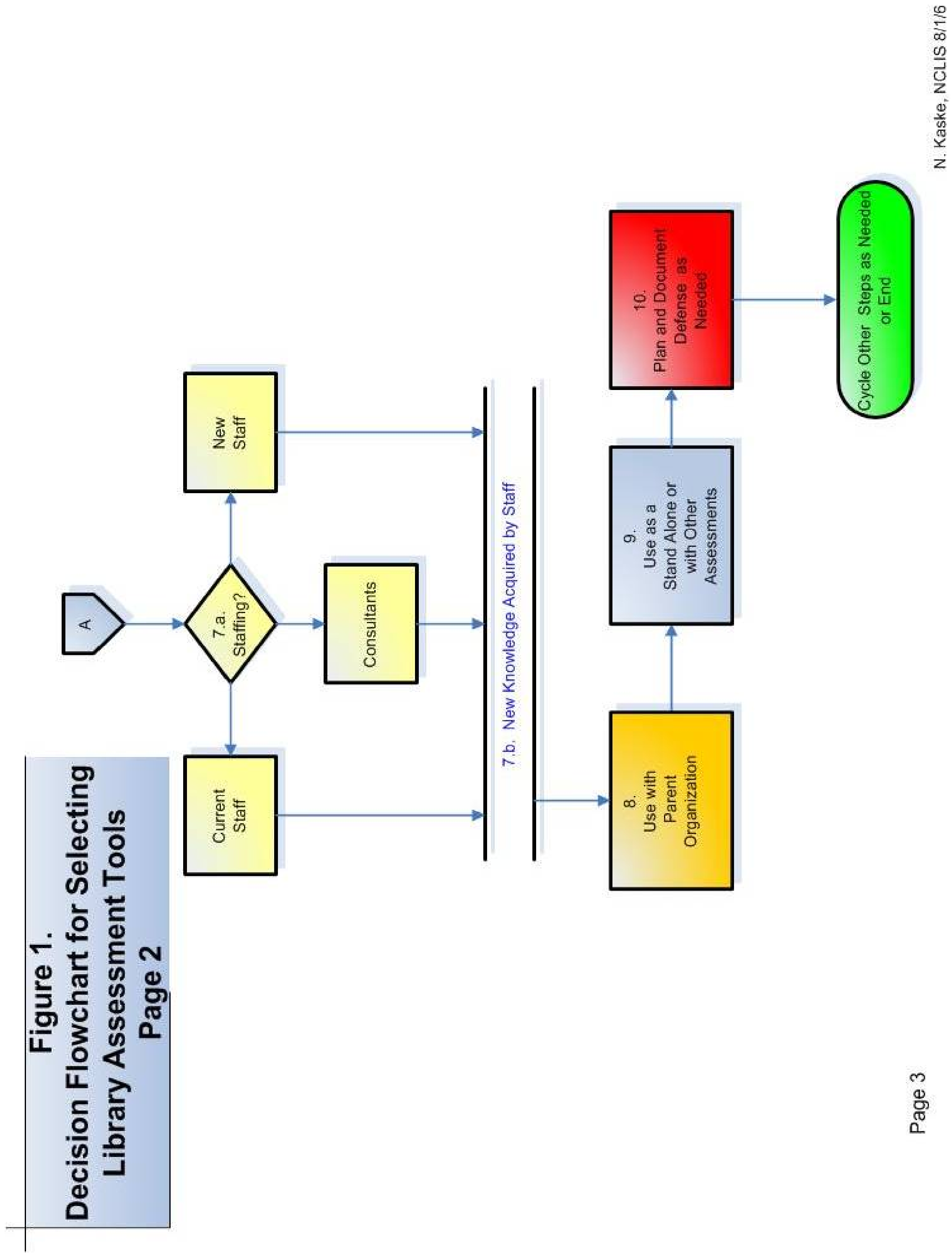
Figure 1.
Decision Flowchart for Selecting
Library Assessment Tools
Page 2
A
7.a.
Staffing?
Current
Staff
Consultants
New
Staff
7.b. New Knowledge Acquired by Staff
8.
Use with
Parent
Organization
9.
Use as a
Stand Alone or
with Other
Assessments
10.
Plan and Document
Defense as
Needed
Cycle Other Steps as Needed
or End
Page 3
N. Kaske, NCLIS 8/1/6

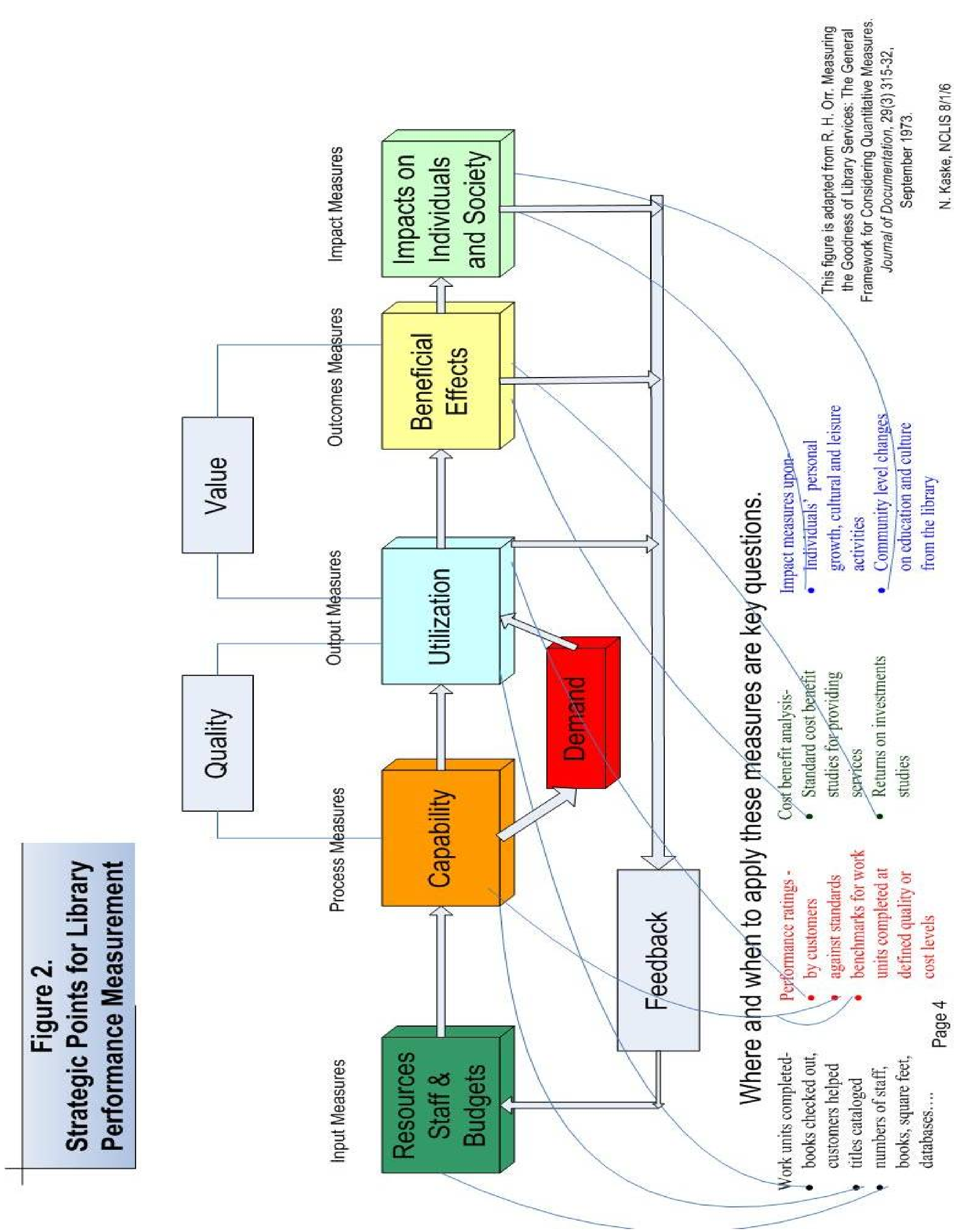

Figure 2.
Strategic Points for Library Performance Measurement
Quality
Value
Input Measures
Process Measures
Output Measures
Outcomes Measures
Impact Measures
Resources Staff & Budgets
Capability
Utilization
Beneficial Effects
Impacts on Individuals and Society
Demand
Feedback
Where and when to apply these measures are key questions.
Work units completed-
books checked out, customers helped
titles cataloged
numbers of staff, books, square feet, databases…
Performance ratings -
by customers
against standards
benchmarks for work units completed at defined quality or cost levels
Cost benefit analysis-
Standard cost benefit studies for providing services
Returns on investments studies
Impact measures upon-
Individuals' personal growth, cultural and leisure activities
Community level changes on education and culture from the library
This figure is adapted from R. H. Orr, Measuring the Goodness of Library Services: The General Framework for Considering Quantitative Measures. Journal of Documentation, 29(3) 315-32, September 1973.
N. Kaske, NCLIS 8/1/6
Page 4

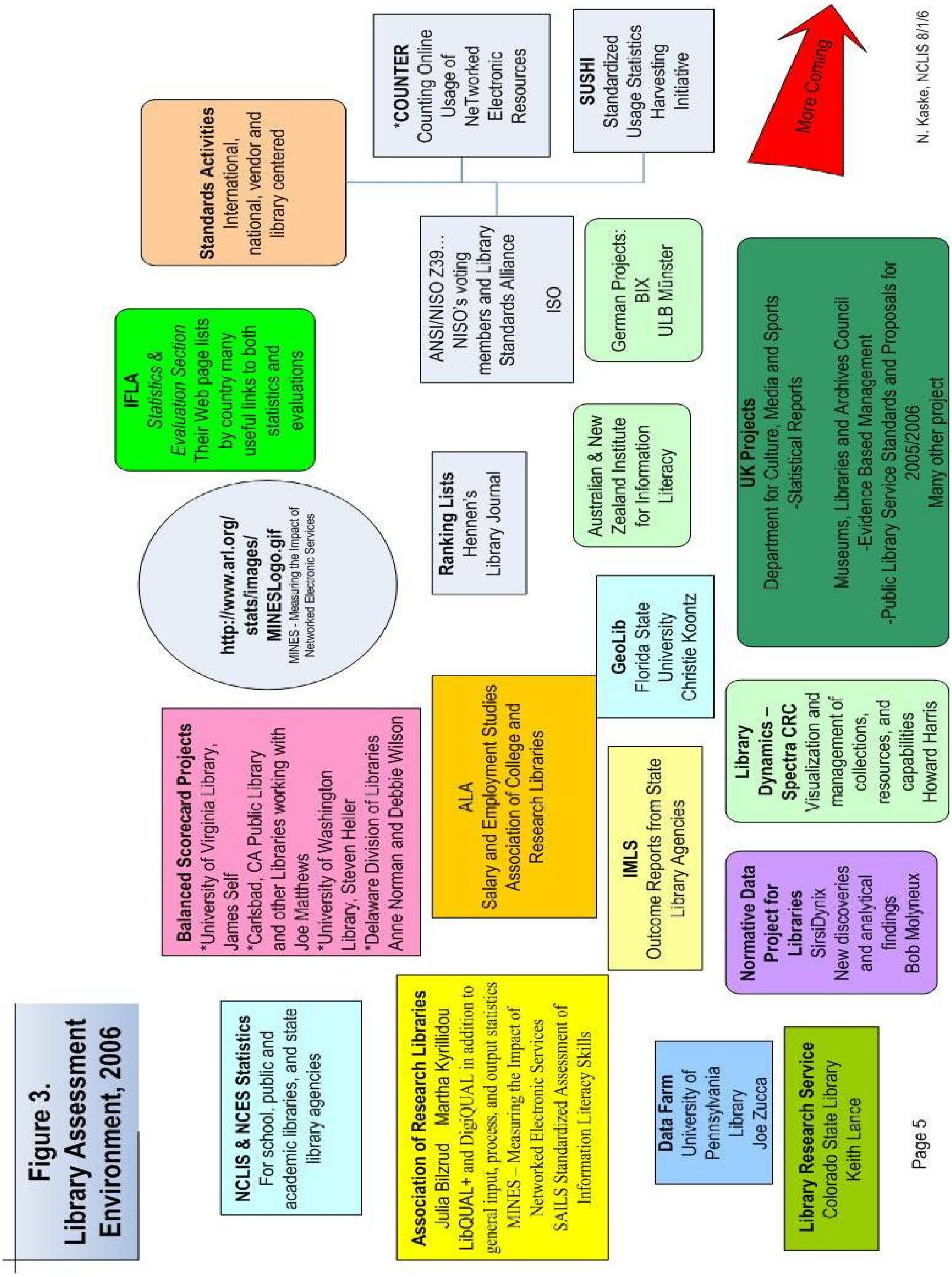
Figure 3.
Library Assessment Environment, 2006
NCLIS & NCES Statistics
For school, public and academic libraries, and state library agencies
Association of Research Libraries
Julia Bilzrud Martha Kyrillidou
LibQUAL+ and DigiQUAL in addition to general input, process, and output statistics
MINES – Measuring the Impact of Networked Electronic Services
SAILS Standardized Assessment of Information Literacy Skills
Data Farm
University of Pennsylvania Library
Joe Zucca
Library Research Service
Colorado State Library
Keith Lance
Page 5
Balanced Scorecard Projects
*University of Virginia Library, James Self
*Carlsbad, CA Public Library and other Libraries working with Joe Matthews
*University of Washington Library, Steven Heller
*Delaware Division of Libraries
Anne Norman and Debbie Wilson
ALA
Salary and Employment Studies
Association of College and Research Libraries
IMLS
Outcome Reports from State Library Agencies
Normative Data Project for Libraries
SirsiDynix
New discoveries and analytical findings
Bob Molyneux
http://www.arl.org/stats/images/MINESLogo.gif
MINES - Measuring the Impact of Networked Electronic Services
GeoLib
Florida State University
Christie Koontz
Library Dynamics – Spectra CRC
Visualization and management of collections, resources, and capabilities
Howard Harris
IFLA
Statistics & Evaluation Section
Their Web page lists by country many useful links to both statistics and evaluations
Ranking Lists
Hennen's
Library Journal
Australian & New Zealand Institute for Information Literacy
UK Projects
Department for Culture, Media and Sports
-Statistical Reports
Museums, Libraries and Archives Council
-Evidence Based Management
-Public Library Service Standards and Proposals for 2005/2006
Many other project
Standards Activities
International, national, vendor and library centered
ANSI/NISO Z39…
NISO's voting members and Library Standards Alliance
ISO
German Projects:
BIX
ULB Münster
*COUNTER
Counting Online Usage of NeTworked Electronic Resources
SUSHI
Standardized Usage Statistics Harvesting Initiative
More Coming
N. Kaske, NCLIS 8/1/6

1. **Why are we evaluating?**

Is it for a one-time go or no-go decision? This form of evaluation is known as a *summative evaluation.* It is conducted at the end of a process. Examples of summative evaluations are: reviewing a librarian for tenure, deciding to discard or retain titles in the collection, deciding whether to select a new integrated online library system or to staying with the current system, or deciding whether to build an additional branch library now or put the decision off for two years.

Is it for an ongoing evaluation done to monitor a service or process with the goal of improving the service or process? If the goal is improvement, then it is a *formative evaluation*. Examples are: ongoing improvement of a library's Web page, fine tuning the collection development plan for higher circulation of items, or increasing the satisfaction level of customers making use of the children's collection.

The differences between summative and formative evaluations are critical. For an in-depth discussion on the differences a number of sources are available from an informal encyclopedia like Wikipedia[i] to a formal text such as Evaluation: Methods for Studying Programs and Policies.[ii]

2) **Who will use the results?**

Is the user internal or external? If the findings will be used as an error check for a process in order to assure high quality standards, then the evaluation may be conducted by staff members. However, if the findings will be used to build a case with the Mayor or the VP for Academic Affairs to justify an additional branch library then data collection and methods for communicating the results will be done differently. In the first case the information on errors will be passed to those who can correct them in an efficient and effective manner. In the second case, the information may be collected and analyzed by a group of library and Mayor or VP for Academic Affairs Office.

The audience and intended use are important factors to consider from the planning prospective and for the means to be used to communicated and hopefully act upon the results. We need to be clear about who will use the findings and if their utilization will be within the library or external. If we do not recognize this, we could have outsiders using seeming low production numbers to assert that the library is an inefficient operation and thus an ineffective organization.

3) **Do we have a baseline or do we need to establish one?**

Baseline for our purpose means a quantity or value known to us and from which we can compare new quantities or values for some task or activity. Examples of baselines are the number of copy cataloged items produced per hour, average number of customer complaints per month, the level of shelving mistakes per 1,000 books made by staff members, and the number of new books stolen per year. Baselines are used to answer questions such as "Did the copy cataloger produce above or below standard last month?" and "Does the library have a 'stolen books' problem from its current circulating collections?" The lack of baseline data prevents us from being able to say we have a productive copy cataloger and we do not have a stolen books problem. To answer the stolen books question we need to know when the books were stolen to determine an annual loss to calculate if a security system would be cost effective. An inventory of only items added to the collection from 1 January 2005 to 31 December 2006 would not inform us about the loss of items added before 1 January 2005 which are a part of the collection. There are cases were the productivity or losses are so great that it is clear to all there is a problem. For example, all the new books on a popular subject are missing and the copy cataloger is only doing 4 titles per day. Normally we establish our own baselines but there are cases where those established by similar libraries or related organization can be used.

4) **What will this tool tell us and what is the precision of its measurement?**

This question is complex. In the first part of the question we need to learn if the tool is valid and reliable. That is, it measures what it is supposed to measure (validity) and it measures it consistently (reliability). We also need to use the tool correctly and for the purpose for which it was intended. If we have a valid and reliable questionnaire that assesses customers' willingness to purchase a given new item or service it may not work to gauge current library customers' willingness to use a new service. Why not? It is because the assessment tool was designed to test peoples' willingness to pay for a new service or product directly, not indirectly as they pay for library services. The cost for the proposed new library service is in our property taxes or tuition. We need to take care to understand and make sure the assumptions for the models and tools we use are applicable to our situation.

The precision part of the question addresses the chose of looking at a sample or the population. When looking at the total population, we do not need to generalize because we have a full census. Samples however have a level and interval of confidence. When we see the results of a poll or study reporting 80% of the people believe a given way with a confidence of 95±5% we know this means if conducted 100 times the results 95 times would be between 75 and 85 (80 ±5 %) and 5 times the results would be greater then 85 or less then 75. Sampling statistics make assumptions about the underlining distribution which also must be acceptable in a given situation. To gain an understanding of these concepts taking a basic statistics class and consulting online statistical labs such as the Rice Virtual Lab in Statistics are recommended.[iii]

The questions, *Do we sample?* and if so, *How large a sample?* must be addressed as well as the question of level and interval of confidence. Question 4 has many aspects which can be summarized by knowing the purpose of the tool, its validity and reliability and its underlining assumptions. Understand and utilizing the findings in accordance with the level and interval of confidence dictated by the sample size and its underlining distribution assumptions are also essential.

5) **What new information will we gain?**

Will this be new information or just newly formatted old information? If the information we are collecting is not new, then we should not be embarking on the effort. Or if the data turn out to be highly correlated to data we are already collecting then we should consider dropping one of the data collections. The cost of data collection and analysis is usually too high to collect redundant data for any length of time beyond verification that the data are redundant. We also should not be embarking on the effort if we do not know how we are going to use this new information in our decision making. The information just being ‘interesting’ is not a good reason to collect, analyze, report, and otherwise support an assessment effort.

The new information should add to our understanding of the current environment or provides insight into library operations. A basic question to answer when determining if we have something new is to ask if the information helps us demonstrate that the library adds value to our parent organization or permits us to manage the library for effectively and efficiently.

6) What are the initial and continuing costs for using this tool?

Initial costs or first-time costs are those of selecting, purchasing, placing, and starting up the use of a new evaluation tool. For example, costs associated with a new system for control and monitoring of database use are initial or first-time costs. This new tool is to help us evaluate use and to permit us to maximize the dollars expended on databases.

Continuing costs or recurring costs are those that accrue from using the new tool. If it makes us more productive–accomplishing more in the same or less time–then it should pay for itself and save money over time. However, if the new tool gives a few additional reports that are of little or no value and require more staff time then it could be a poor decision for one key reason. The costs for the new tool are greater then the value of the additional information and useful knowledge gained. A word of caution, in many cases the initial cost are low and the possible saving from use look promising but there are many elements of the continuing costs we do not see at first. Take care to clearly understand the total cost of your new evaluation tool.

The classic *initial vs. continuing costs* problem can be expressed by the decision to purchase a hybrid vs. a standard gasoline fueled automobile. Generally, the hybrid car has a higher initial cost but lower continuing costs. However, this is all subject to the assumptions made. These would be the expected life of the car, how long it would be retained, fuel and repair costs, and the expected value of the car when sold or traded. These assumptions are fundamental elements and should always be expressed explicitly. This is especially true in our highly computerized library environments with the complex human-computer interaction to consider. For example the amount of staff time to be saved turns may turn out to be a false assumption. When implemented the new efficient process has dumbed down the job, resulting in lower morale and productivity. We should remember to test assumptions with all staff affected to limit unattended consequences.

7) What are the staffing requirements and what new operational knowledge does the staff learn?

To answer these questions we will need to ask some additional questions.

Staffing requirement:

- Will the new assessment tool require staff training?
- Will current staff be able to accommodate the additional workload or will overtime or temporary staff be required?
- Will specialized staff such as a consultant be required?

If the work is specialized and requires a consultant then that person will require selection and supervision as will the temporary staff.

Operational knowledge acquired by staff

- Will the new knowledge help staff provide better service or a capability to make more informed decisions?
- Will the library organization be able to retain and utilize this additional knowledge if temporary staff or a consultant is used?
- How will the new operational knowledge be transferred to permanent staff?

Given that library organizations must learn if they are to survive in our ever changing environment, the question of who gains this new operational knowledge and how it is shared in the library becomes critical.

8) Will the assessment resonate with and help support the goals of the library's parent organization?

So, what has the library and its staff done for us lately? The answers to these questions are critical for the vast majority of libraries as few libraries stand as independent organizations. The city, county, school, cooperative, or university administers who oversees the library needs this question answered in a clearly understandable way. These administers need to see how the library adds to the city's economy, the educational achievements of the students or to the company's bottom line. The individuals that use the library and its services may benefit from this information. Having customers aware the library is providing effective and efficient services will clearly help in demonstrating the library organizations and their staffs add value to the parent organization, be that a city, academic, or business establishment. Accountability is increasingly the norm. Just because people love books and libraries is no reason to fund libraries. The administers must see the added value coming from the library's collections and services.

If the information from the new evaluation tool does not resonate with the parent organization then we must ask, is the tool needed for internal use? If the answer to this question is a clear yes and not "it would be nice and the information would be interesting to know" then it may be a reasonable addition. Be careful of redundant measures and ones of only passing interest. Remember resources are finite.

9) How will the findings be used by the library's parent organization?

Packaging the message from the assessment is critical. *Is this assessment a one off or is it a part of a set? Can the parent organization incorporate the library's assessment as part of its budget or 'legislative' case?* If yes, then it is a useful assessment. But if no, your assessment may be seen as superficial, unscientific, or just library *mumbo jumbo*. Demonstrating the value of a library to the parent organization is a critical step because administrators must explain and understand the value added by library collections and services to their board of directors and their customer base. Some of the most effective library advocates are school principals, college presidents, city managers, mayors, and CEO's of business organization.

10) How might the findings from the assessment be used *against* the library?

It will not matter if your measure was for *internal use only* if it is used by someone to make the library appear poorly managed or ridiculous. To prevent this, one needs to take care in the methods of recording and reporting assessment data. Secrecy is not being suggested, quite the contrary. Evaluation should be open as long as individual workers rights are protected. One needs to create, build and maintain a culture of assessment. This is no easy task and is the subject of many articles and books. Findings should be presented in relation to past performance and established strategic goals and objectives for the library and in some cases the parent organization too. For example an assessment of the library's information literacy programs shows that professor are requesting 35% more classes is year over last year. Expanding this effective program will require additional funding. Demonstrating cause and effect is difficult but reasonable correlations, well articulated, suggest the need for additional funding. This is where the art and science of library assessment come into play with the art and science of management. A key to success is making decisions based on the best qualitative and quantitative information available by using reliable evaluation tools. Without solid information it is just BS–boastful statements.

Final Comments

The questions offered in this paper about the choosing of the best evaluation tools for library assessments are numerous and sometime complex. However, once answered the tasks related to evaluation and the use of the findings should help as us all demonstrate that library adds value to our parent organizations and society as a whole.

---

[i] http://www.wikipedia.org
[ii] Weiss, Carol. H. Evaluation: Methods for Studying Programs and Policies, 2nd edition, Upper Saddle River, New Jersey: Prentice Hall,1998. ISBN 0-13-309725-0
[iii] http://www.ruf.rice.edu/~lane/rvls.html