Gathering and Compiling Library Statistics for Evidence-based Decisions

LALINC Statistics Task Force

By PresenterMedia.com
LALINC Statistics Task Force

Background: Gathering and compiling statistics for library resources is a conundrum for many within the consortium. Questions arise on how to best gather and compile data in order to respond to administrators, accrediting agencies, affiliated organizations, etc. Responding to the questions, much less establishing best practices, is a challenge due to the plethora of tools and data available.

Charge: The STF is tasked with reviewing the issues, revealing common needs and identifying next steps towards enabling consortium members to address those needs.
LALINC Statistics Task Force

MEMBERS

Lora Amsberryaugier, University of New Orleans
Patricia Brown, Louisiana State University at Eunice
Shanna Clevenger, Delgado Community College
Alice Daugherty (Chair), Louisiana State University
Angela Dunnington, Southeastern Louisiana University
Kevin Hebert, Xavier University of Louisiana
James Hodges, University of New Orleans
Margaret Keller, Northshore Technical Community College
Aaron Lercher, Louisiana State University

Megan Lowe, University of Louisiana at Monroe
Karen Niemla, University of Louisiana at Monroe
Jerome Marcantel, McNeese State University
Sarah Simms, Nicholls State University
Lori Smith, Southeastern Louisiana University
Mike Waugh, Louisiana State University
The LALINC Statistics Task Force is hoping to share our problems and our solutions!

What data are we collecting?

Are we all counting things the same?

To whom are we reporting our statistics?

Can we present value, ROI, or assessment measures with our numbers?

Do we tell a story?
Three Stages in a Schematic Workflow

Stage 1: Validity of the source data:

Stage 2: Validity of data processing:

Stage 3: Validity of decisions:
Three Stages in a Schematic Workflow

Stage 1

Validity of the source data:
- What is the data we are working on?
- Is it correctly categorized?
- Is it counting what it’s supposed to count?
- Does it come from the right places?
Three Stages in a Schematic Workflow

Validity of data processing:
- What is done with this data?
- How do we ensure that we don’t lose the initial validity of data?
- How do we label it and arrange it to make it usable?
- Do we have the right data and the right comparisons?
- What other data is relevant?
Validity of decisions:

- What is the data used for?
- How do we report our results?
- How do we ensure that data is not misinterpreted by others?
- What is the data being used for?
- How do we assemble data to help decision-makers avoid foreseeable mistakes?
WHAT IS A “SEARCH?” EXAMPLE

EBSCO searches in the Interface Usage Report
One search counted for each database in the search.

80 databases = 80 searches per query

“Keep in mind that due to the nature of statistical reporting, the Interface Usage Report is not a true indicator of exactly how many searches your users conducted.”

Also counts a search with subsequent filtering or refinements

If you need a search number that counts the number of queries performed by the user, try using the count of the most searched individual database.
VOLUME OR ITEM EXAMPLE

Numbered collections of journals or books
Usually, an item count will suffice
But if there are duplicates, items and volumes may be different numbers
Downloads versus Page Views/ Counter versus non counter

COUNTER Book Report 2 counts “successful section requests”

Most ebook vendors equate successful section request as a chapter download

eBrary uses pages viewed for 10 seconds, copies made, pages printed, instances of PDF downloads
Double counting?

How do vendors count HTML and PDF downloads?

The COUNTER standard eliminates double counts if the same article in the same format is accessed multiple times during a short time frame.

However, if the same articles are accessed via HTML first and then PDF within a short time frame both formats are counted.
Mapping payment information to usage information can be tricky because of backfiles and open access.

For example:
Sometimes there is one title with payment information, but several same titles with usage.

*Journal of Geophysical Research: Atmospheres $xxxx.xx*

Which title for usage?
*Journal of Geophysical Research: Atmospheres*
WHY DO WE PROCESS OUR DATA

Don’t lose validity of source data
- Save each step of process
- Manage files
- Make mistakes
- Fix mistakes

Do make it possible to use data for decision-making
- Don’t distract your supervisor with unneeded data
- Don’t mislead your supervisor with false impression of complete data
- Don’t mislead your supervisor with spurious relationships among data
DATABASE DESIGN

What objects are the data about?

• What counts as an object depends on the purposes or on what decisions are about
• For librarians, answers often based on lots of prior work defining bibliographic objects.
  For example, FRBR entities: Work, Expression, Manifestation, Item
• Good to use prior work
• Bad when entities don’t fit purposes

Each table (=relation in a database) represents a kind of entity

Distinct objects in each row of a table

• ISSN objects in a publisher’s price list
• For a particular year, a JCR SERIAL-YEAR objects in Journal Citation Reports
• JOURNAL or JOURNAL-PART objects are different from previous two.
  Rows in the table of Cost per Citation calculations at Bergstrom’s and McAfee’s www.journalprice.com
  Or perhaps some other entity, described elsewhere

Relationships between entities

• Each ISSN has one JCR SERIAL-YEAR, but there may be many ISSNs for one JCR SERIAL-YEAR
  Electronic and Print
• Similarly each JOURNAL has one JCR SERIAL-YEAR (for that year), but there may be many JOURNAL-PARTs for one JCR SERIAL-YEAR
Database Design problem

JCR Serial-Years

1:N

Price list ISSNs

1:N

N:M

Journals or Journal-parts

Bad and/or confusing
Database Design solution 1

JCR Serial-Years

1:N

Price list ISSNs

N:M

Journals or Journal-parts

1:N

Set of pairs of ISSNs and Journal-Parts (aka “Intersection Relation”)
**Database Design solution 2**

- **JCR Serial-Years**
- **Expanded Price list ISSNs**
- **Journals or Journal-parts**

Add enough ISSNs so there's at least one for every JCR entity and one for every Journal-part.
SOURCES OF CONFUSION

1. Entities (=Tables=Relations) are defined by others for their purposes.
   - Publisher’s price list
   - JCR
   - ISSN system

2. The confusion generated by multiple ISSN for each JCR Serial-Year, and by multiple Pricelist-Journals, is not prevalent.
   - It’s a waste of time to work at entirely resolving it
   - But it’ll come up
ADDITIONAL DATA ISSUES

File naming, File management, Data Cleaning

• Internal – how the file is labeled for file management
• External – how the file is labeled for others

• Without naming convention:
  • Meeting minutes. Docx

• With naming convention:
  • 20141013_ACRLTaskForce_Minutes_v1.docx
ADDITIONAL QUESTIONS?
THANK YOU FOR JOINING US TODAY!

LA LINC Statistics Task Force

Need further information?
Contact: Alice Daugherty (Chair)
adaugher@lsu.edu

Make sure to follow the Statistics Task Force beta wiki!
Site created and hosted by Karen Niemla,
LSTF wiki: niemla.info/lstfwiki/