Automating Cornell University Library's (CUL's) annual data collection using LibPAS

2012 Library Assessment Conference

Linda Miller (LM15@cornell.edu) and Joanne Leary (JL41@cornell.edu), Assessment & Communication, Cornell University Library

A prefatory glossary:

- Collection: A collection of measures.
- Group: A grouping of measures within a collection.
- Location: An entity at which data is reported (a unit library, a unit library subdivision, an affiliated group of libraries, a campus, etc.).
- "Rolling-up" data: In LibPAS, locations are arranged in a hierarchy. Most measures are summed up the hierarchy, creating totals at each location, including all of the hierarchically higher locations.
- Exception value: A pre-defined value that can be selected when a response is not applicable or available (e.g., unavailable, not applicable, reported jointly).
- Input forms: A type of collection that populates data into another collection.

Opportunities:

- Customizable, decentralized reporting. Access data from any computer, with no limit on the number of user seats.
- Chance to review and rethink everything, involving others:
  - Review which measures are collected, and which dropped.
  - Rethink reporting: create more audience-specific reports; provide easier & more flexible access to data notes.
  - Review, simplify and better document processes and policies.
- Build goodwill through simplifying collection processes and by expanding access to data.

Challenges at CUL:

- Some locations report together for some, but not all measures.
- Some data are reported across many locations.
- The storage facility is both a unit and a shelving location.
- Frequent measure and organizational changes.
  - A few libraries have consolidated; some have become "virtual;" some central units have frequently been moved in the organizational structure; etc.
- Nationally requested, breakout expenditures measures are reported jointly for endowed units, but separately for state units.
- A few measures that are generally only collected at the CUL level are requested for one unit at the unit-specific level nationally.
- Different library management systems are used at the Ithaca/Geneva and NYC campuses.
- Confidentiality issues with expenditures reporting for small units.

Conceptual shifts:

- Data is available to all trained users, at all times, from any computer. Data is available pre-publication.
- More emphasis is needed on usability for end users.
- Greater opportunity and need to focus on the consistency of data elements over time.

Decision points:

- **Preliminaries:**
  - Who is responsible for implementing and maintaining the system and for training staff? Who is backup?
  - How should the hierarchy of reporting locations be set up?
    - Which location groupings are frequently needed?
    - Do locations report jointly for some measures?
    - Is there any data that cuts across many locations?
    - Is there a storage facility?
    - Are there central units frequently moved in the organizational structure?
  - How often will data be collected, and at what location level?
  - Which measures should be included? Which dropped?
  - Will any measures not collected centrally be included?
  - How many retrospective years will be input?
- **Collections:**
  - How many collections need to be set up? Is there any data that should not be available to all staff?
  - How will the measures within a collection be grouped, ordered and named?
  - Should all measures be summed during "roll-up"?
  - What exception values are needed?
  - How will definitions be input and updated?
  - Where should overarching data notes be recorded?
- **User access / Collecting data:**
  - Who should have access to the collections?
  - How many data input forms are needed?
  - How will data collected centrally from online systems be added?
  - Will exception values be entered for locations not reporting values for particular measures?
  - Will corrections be made in previous years' data, or just made in data notes?
- **Reporting / Training:**
  - How should staff be trained for inputting and reporting?
  - How should it be indicated that data is still in draft?
  - What documentation is needed to aid report writers?
  - How do users at a location best share their dashboard?
  - Which report templates should be made available to users?
  - Which reports should be published?

Recommendations:

- **Preliminaries / Generalities:**
  - Start well in advance of the target implementation date. Allow time for usability testing.
  - Have multiple people responsible for the system, with one having primary responsibility.
  - Consider all set up issues seriously. However, don't let any hard decisions delay you unnecessarily; the system is very flexible.
  - Take the time to involve managers across the Library to decide which measures should be included and which dropped.
    - Exceptions require extra measures (e.g., the storage facility as a unit and a as a shelving space). Make sure the exceptions are needed.
  - Decide up front how many retrospective years will be added, and inventory location and measure changes over that time frame.
  - Have at least one retrospective year input when you do your first data collection to showcase systems capabilities and gain experience.
  - Learn all key system functionality up front (on a smaller scale) to aid in set up decisions. Create a test collection.
  - Avoid duplicating work outside of the system.
- **Location set up:**
  - When creating the location hierarchy, consider how data will be reported out most often. Create additional levels when units share multiple measures; otherwise use exception values. Balance power with ease of use.
  - Consider a cross-unit location rather than reporting at both "roll-up" and subordinate locations.
  - Consider keeping central units that are frequently moved in the organizational structure as separate reporting locations.
  - (Note that you may need to adjust the name of any unit reporting separately nationally to be represented properly in LibPAS's national collections.)
- **Collections set up:**
  - Have a separate collection for any confidential data and virtually reflect any appropriate summary figures to a collection staff have access to.
  - Keep naming and measure order conventions as consistent, and as helpful as possible for your users. CUL:
    - Puts calculations at the bottom of measure groups, with the measure name for the total being the shortest.
    - Locates alternative totals after the final totals, in separate measure groups.
    - Includes start and end dates in measure names.
    - Includes "(calc)" at the end of measure names for calculations.
    - Uses "cannot be categorized" to handle differing levels of measure breakouts over time.
    - Uses group names to guide users when helpful (e.g., if no longer collected; if subsets of other measures)
  - Consider using the powerful copy measure function when building collections.
  - Space out the sequence numbers assigned to measures within groups to allow measures to be inserted more easily at a later date.
  - Define a central set of exception values, and differ only as needed.

- **Data collection:**
  - Review with unit heads who should have access to confidential data. Test sign-ons as set up.
  - If users provide data for multiple locations, create them separate but similar sign-ons. Don't give input access to more locations than reported for.
  - Input forms should indicate which year they are currently for as measures collected can change from year to year.
  - Create enough input forms so that units reporting for a much smaller number of measures don't have to wade through them all. However, feel more comfortable creating fewer input forms as data providers can see what was entered in the previous year.
  - Import data collected centrally through online systems yourself, saving import files from year-to-year to copy for re-use. Have all units include a count or exception value for each measure.
- **User reporting / Training:**
  - Train staff before giving them reporting rights.
  - Include "DRAFT" in the title of the year still being collected.
  - Create an accompanying online portal to support both data providers and users.
  - Create cheat sheets and FAQs, and provide easy access to official annual data sets to aid reporters.
  - Consider making available an Excel file of data notes that can be filtered across years, locations and measure groups; include an "overarching" data note category.
  - Consider indicating which years canned reports are good for.
  - While reporting is much easier, assume you will still do the lion's share.
- **Official annual reporting:**
  - Report separately for external and internal audiences. Easily create unit-specific reports for Library staff.
- **Maintenance:**
  - Start preparing for each new annual data collection well in advance.
  - When possible, time data roll-ups so that sums at higher locations are complete, or obviously incomplete.
  - Put data notes about corrections in the year they apply to.
  - Keep a longitudinal report that provides top level sums for all key measures, so you can easily check that changes were made correctly.
  - Track changes in definitions over time in the system.

Where we are now:

- Quick customizable reporting, from any computer, with 5 years of data in LibPAS.
- Data providers happier. Offering training for interested staff for ad hoc reporting annually.
- Multiple reports created for different audiences: longitudinal Excel report (at the CUL level) for external audiences, and unit-specific LibPAS trends and percentage change reports at the online portal for Library staff.