Figuring the True Cost of e-Products: e-journals in an e-package world

Ted Koppel

Verde Product Manager, Ex Libris
!! Warning !!

STATISTICS AND CALCULATIONS AHEAD
What I will not do in this session

- Solve pricing issues for libraries
- Convince content publishers to change their ways
- Lobby for or against “big deals” or similar
- Make decisions for libraries
- Take a stand on the validity of cost-per-use as an indicator

I will suggest ways of calculating the actual cost of e-products based on different approaches to assigning value.
Why talk about this at all?

- Right or wrong (and it is being debated), Cost Per Use calculations are being used as a factor in library retention decisions.

- Usage information in the SUSHI and post-SUSHI era will be easier to retrieve and be far more standard in structure and scope than ever before.
Why talk about this at all (2)?

ERM systems (from all vendors) are moving into their second or third versions and, with SUSHI, are mature enough to take on the calculations required.

But … because most purchasing is done at a package level, accurate e-journal (e-constituent) level cost information has been very hard to discern.
A short detour to ERM Basics
DLF-ERMI Specification

- Called for e-product management functionality at three levels:
  - Interface or platform
  - Package
  - E-constituent or e-journal

- Most (perhaps as much as 85%) of e-product management in ERM systems takes place at the package level
Package types

- **Aggregator packages**: library customer buys the package as-is; no ability to pick and choose. (Similar to the way that the cable company works)

- **Selective packages**: library can pick which titles in a package are to be included in their subscription
Selective package types

- **Selective complete** (functionally, acts like an aggregator package)

- **Selective incomplete** (pick and choose)
  More complex functionality for ERMs

Selective package pricing models are
- Creative and idiosyncratic
- Inconsistent
- Complex
- Sometimes inscrutable
Why now?

- SUSHI and COUNTER provide an easier/better way of collecting usage statistics

- ERMs should do the following tasks (among many others):
  - Collect and store usage data
  - Collect and store Acquisitions and Invoice data (either looked up, locally edited, or transferred from ILS)
The real issue

- Getting from package-based pricing numbers
- E-journal based costs
Assumptions / Illustrations

- Package has 5000 e-journals
- List price for my institution = $18,500
- Discounted (actual) price = $15,000
- Invoice amount = $15,000
- Broad-spectrum index/full text provider
Approach #1: Publisher supplied

- E-content Publisher (or perhaps agent) makes available per-journal cost to library

- Based on what? List price? Invoice amount?
- Available at all?
Approach #2: Straight Line Division

- Easy to figure:
  - $15000 ÷ 5000 titles = $3 cost per title

- Seems low, inaccurate, and awfully unfair

- All journals are not created equal
Approach #3: Unique title calculation

- **Rationale**: You’re paying for many titles more than once. Figure out what percentage of this package’s titles are **unique to this package**, and then do the division. Most ERMs can create unique and overlap lists.

- $15000 ÷ 2000 unique titles = $7.50 per unique title.
Approach #3: But wait a moment

- Is calculating on the basis of unique titles fair? Isn’t there some value in the titles that this package delivers, even if they are found elsewhere?

- Even if uniqueness is a reasonable criterion, not all titles in that unique subset are of equal value
Approach #4: External, objective quality assessments

- Use ISI Impact Factor, PageRank, or lists of “most important journals in <field>” as objective third-party sources

- Factor in interface usability

- Rationale: An authoritative “Someone else” determines what’s of high quality; the library uses those decisions to identify titles and calculate costs
Approach #4 - disadvantages

- ISI Impact Factors are discipline based; an IF 1.05 in discipline A isn’t really equivalent to an IF 1.05 in discipline B.

- Impact Factors not available/calculated for all journals/disciplines.

- ISI themselves have issued various warnings and advisories about impact factors and why they should not be used as a single determining factor.
Approach #4: calculations

Assume that library decides:
- Physics: any IF greater than 1.652 means quality
- Biology: any IF greater than .857 means quality
- And so on for each discipline

Of the 5000 original titles, 400 are considered quality publications based on the IF scores meeting or exceeding the library’s threshold

$15000 \div 400 \text{ quality titles} = \$37.50 \text{ / title}$
Approach #4 – But wait, there’s more!

- With a little more work, we can make use of the IF, turn it into a multiplier, and do some real ranking within a discipline

- Example: 50 of the titles in the quality group of 400 were in Physics. (that’s 1/8)

- $15000 ÷ 8 = $1875 represented by Physics
**Approach #4: calculations**

- Add the values of all the impact factors, convert them into percentages for weighting.
- Multiply (percentage per title * $1875) to arrive at the cost per high quality physics title.

**Example:**
- Add all Physics titles IF = 34.40
- Annals of Physics has IF 3.20
- $34.40 \div 3.20 = 10.75\%$ of combined IF
- $1875 \times 10.75 = $201.56$ Annals of Physics weighted cost based on impact factor calculation.
Approach #5: Usage based

- **Rationale:** Assumes that value to library of electronic resource is represented by usage.

- **Flaws:** (a) Usage doesn’t always mean quality – in fact, it may mean the opposite. (b) Is past performance indicative of future value?

- **How?** ERM systems do/will have the ability to collect usage data (through SUSHI/COUNTER)
Approach #5 – Usage based calc

- Libraries can decide for themselves what level of usage is worth counting. Test the 80/20 rule.

- Example:
  - Of our 5000 titles, only 1000 have usage > 5 last year. Therefore, 1000 titles = $15,000 cost
    - Title m1 – 98 uses – 1.09% of total = $163.50
    - Title m2 – 74 uses -- .77% of total = $115.50
    - Title m3 – 50 uses -- .53% of total = $79.50
    - And so on
Approach #5 – Usage Based Calc

Considerations:
- Based on local, not external third party usage
- Depending on where lower threshold is set, can count incidental or accidental uses as real usage
- Calculation heavy
Dependencies with all approaches

- Relies on accuracy and granularity of incoming COUNTER statistics
- Relies on local ERM to store and manipulate usage and financial data
- Complexity for library to understand and identify their approaches and applicable thresholds
- Not all publishers support SUSHI today – how many will next year? 2009?
But there are benefits to knowing your costs

- Cost-per-usage as an indicator for purchasing and renewal decisions

- *May* lead more transparency – truth in pricing – by publishers

- *May eventually* lead to breakup of the package model as libraries become more savvy about their real costs
Big issues still to be addressed

- Open access and related pricing models – is there value in something that you don’t pay for? How do you show it?

- “Big Deal” pricing (and particularly Big Deal for consortia) – when is a selection not a selection? When is a purchase not a purchase? Is access an implicit purchase?
In Conclusion

- E-product purchase and use continues (and will continue) to rise

- Librarians need to show good custodianship of institutional funds – one success indicator being use

- As a profession, we need to determine how best to measure usage and electronic product quality, just as we have always done for books
Questions?

Ted Koppel
Verde Product Manager

ted@exlibrisgroup.com