2015-02-24 breakout: Usage Data

February 24 LD4L Workshop breakout session: Usage Data

facilitator: Paul Deschner

Usage data sources

- OCR-ed bibliographies and page rank
- ILL usage
- Yahoo circ logs
- Web analytics (e.g., DPLA UI analytics, esp. contextual granularity)
- Search terms as form of usage; also as compared to other usage data
- Entities extracted from queries, not simply literal queries themselves
- How often a link is traversed; how many times your link has been reconciled in triple store
- Browsed materials
- Citations; also citation networks as compared to other usage data
- Course-book lists across institutions

StackScore

- Makes data muddy
- Too many metrics mixed together; need to separate out the metrics
- Common metrics needed across institutions
- Computational transparency important: metrics and algorithms

Use cases

- Keeping tabs on popularity of colleagues' publications
- Usage data as diagnostic tool for targeted collections: highly invested-in parts of collection not being used could drive arranging an exhibition to increase awareness
- Scholars doing research on other scholars' research and publications
- Look at when items were used: what was checked out in last week, month, year, etc.
- Link traversals and other link metrics could be sent to link's publisher

Privacy

- Opt-in option for users willing to share their usage data
- Huddersfield University (England): more liberal approach to data exposure, including access to clustering (users who borrowed this also borrowed that) and usage by academic course and school
- IP-based web stats inherently less risky than personal ID-based circulation data
- Anonymization tools important
- Clustering dangerous

Long tail issue generally and at own institution

- Options: random selection out of tail for exposure, subject-filtered selection
- Important that UI expose long-tail possibilities prominently, above the page-fold
- Usage data from other institutions and ILL balances out local institution's biases

Negative usage data at local institution

- Important to see what users are looking for but local institution doesn't have
- What doesn't circulate in-house but is available via ILL
- What isn't read at Columbia but at Yale

Usage data runs risk of becoming prescriptive

- Blandness of collections when everyone acquires most popular items