Comparing CONTENTdm Data to Google Analytics Data

Chris Schommer
Special Projects Associate, Macalester College

Johan Oberg
Digital Scholarship and Services Librarian, Macalester College

All logs/total activity

- All Logs
- Item Viewer Only
- "Bots removed"
Why the large discrepancy?

• IP restricted vs open access vs password protected
• Data logs include every server transaction
  – Diagnostics
  – User interface images
  – Administration etc
• This test excluded 15 collections, which made the results questionable

Data overload

• Our server logged 600,000 data transactions
• 100,000-600,000 lines of information a day
• Visual double checking of your search results is nearly impossible
• The more specialized the search, the less confidence I had in the results
Adding Google Analytics to CONTENTdm

“CDM4 > Includes” folder

Insert the Google Analytics script into “global_footer.php”

To limit the number of tracked collections, create a custom footer for each collection, by copying “global_footer.php” - renaming it “collectionalias_footer.php” – and insert the script here.
How Google Analytics Manages Data

Focus on the visitor or the person.

Easy to set date range.
How Google Analytics Manages Data

Data does not seem to contain visits from bots or crawlers: 167 off-campus pageviews from Minneapolis/St Paul.

Easy to explore different aspects (e.g. direct traffic vs referrals vs search engines, geography, browsers, languages, returning vs new visitors, etc.)
How Google Analytics Manages Data

Google Analytics captures keywords from search engine traffic. Potential uses:

Metadata modification? “Recent searches”?

Data Privacy?

- CONTENTdm logs all IP address
- Google does (we think), but is not readily available to the Analytics user.
- Users can now opt out of Google Analytics with a browser plugin:
  http://tools.google.com/dlpage/gaoptout
Questions