Measuring Repository Use at Texas A&M University

Anna J Dabrowski

Abstract

This poster describes how usage data are currently gathered for material within the OAKTrust repository. It focuses on file downloads, shows the scope and limitations of data gathering tools, and identifies directions for improvement.

Rationale

Usage data help to quantify the impact and value of institutional repositories (IRs). File downloads over time are a metric indicating user interest, and intent to interact with IR material.

•

Despite having multiple tools to track downloads, robot (bot) traffic and different methods for counting downloads make accurate and comprehensive data difficult to acquire and display to users.

•

Multiple tools and institutional resources can be combined to improve usage data.

Tool	Metric	Scope
DSpace	"Bitstream/File visits"	• Built-in log analysis tracks visits to each file in an item. • Available to users on item "Statistics" pages. • Concern: Poor bot filtering.
DSpace Stats Add-on	"DSpace downloads"	• Weekly scrape of Dspace item "Statistics" pages aggregates all file visits by item. • Available to users in widget on item pages. • Concern: Lag; Poor bot filtering.
Google Analytics	Download "Events"	• Code on repository pages tracks file visits. • Available to administrators through login. • Concern: Limited to clicks from within repository.
RAMP	"Citable content downloads"	• Automatically gathers and stores Google Search Console data. Allows for comparison to other institutions. • Available to administrators through login. • Concern: Limited to downloads from Google Search properties.

Future Directions

- Regularly update list of known bots.
- Contribute to updating existing add-ons or developing new statistics features.
- Currently also gathers logs, collaborate to include analysis and additional logic for identifying bots.
- Extract, consolidate, and display download counts from Google Analytics "Events" and the Google Search Console.

References

Annex J: List of internet robots, crawlers and spiders. (2016). Counter Project. Retrieved from https://www.projectcounter.org/code-of-practice/appendices/850-2/

Carvalho, J. (2010). StatisticsAddOn. DuraSpace Wiki. Retrieved from https://wiki.duraspace.org/display/DSPACE/StatisticsAddOn

Diggory, M. and Lawrence, A. (2016). SOLR statistics. DuraSpace Wiki. Retrieved from https://wiki.duraspace.org/display/DSDOC5x/SOLR+Statistics

Greene, J. (2016). Web robot detection in scholarly Open Access institutional repositories. Library Hi Tech, 34(3), 500–520. doi:10.1108/LHT-04-2016-0048

OBrien, P., Arlitsch, K., Mixter, J., Wheeler, J. and Sterman, L. B. (2017). RAMP – the Repository Analytics and Metrics Portal: A prototype web service that accurately counts item downloads from institutional repositories, Library Hi Tech, 35(1), 144–158. doi:10.1108/LHT-11-2016-0122

OBrien, P., Arlitsch, K., Sterman, L., Mixter, J., Wheeler, J. and Borda S. (2016). Undercounting file downloads from institutional repositories. Journal of Library Administration, 56(7), 854–874. doi:10.1080/01930826.2016.1216224

Walsh, M. (2016). DCAT meeting August 2016. DuraSpace Wiki. Retrieved from https://wiki.duraspace.org/display/cmtygp/DCAT+Meeting+August+2016#DCATMeetingAugust2016-HistoryofDSpacestatistics

Poster available at: http://hdl.handle.net/1969.1/160326

OAKTrust Open Access to Knowledge

LIBRARIES TEXAS A&M UNIVERSITY