

“Plagiarism is Good”

Moving from Access to Use as Metrics for
OCW/OER Use and Reuse

Brandon Muramatsu, mura@mit.edu
Tom Caswell, [@tom4cam](mailto:caswell.tom@gmail.com)
Flora McMartin, flora.mcmartin@gmail.com

November 2010

Citation: Muramatsu, B., Caswell, T., McMartin, F. (2010). Plagiarism is Good: Moving from Access to Use as Metrics for OCW/OER Use and Reuse. Presented at OpenEd 2010: Barcelona, Spain, November 3, 2010.

Unless otherwise specified this work is licensed under a Creative Commons Attribution 3.0 United States License (creativecommons.org/licenses/by/3.0/us/)

About the title: “Plagiarism is Good”

We’re not really talking about plagiarism—but rather use within the terms of the license of most OCWs/OERs...to use, as well as create modifications and derivative works.

We want to see others using these materials, in their own work or on their sites.

We fully support and encourage the proper attribution of the author(s) as required by all Creative Commons licenses.

Others, here at OpenEd 2010

- ▣ **Web Scale Search / Nathan Yergler**

- ▣ How do we measure efficacy of web scale search?
- ▣ Feedback loop...What were positive experiences of search?

- ▣ **Learner Analytics / Erik Duval**

...we're interested in moving these to the forefront...

The 30 second preview

- “Plagiarism is good” is an idea we’ve been talking about for 3 years now
- Move beyond access metrics, to “use” metrics that many sites are capable of evaluating
 - Access metrics are “easy”, and can be “standardized”
 - But, they’re like McDonald’s—“Billions Served”
- So, let’s start using something else...
- Can we, collectively, identify a more interesting set of metrics and questions, and what might we do to answer them?

We’re still looking for the time to really work on these ideas...

The “early” days...

- ❑ **First started talking about this idea about 3 years ago...**
 - ❑ Based on 12 years of work with educational digital libraries in the United States (1995-circa 2007)
 - ❑ NEEDS, SMETE.ORG, MERLOT (~9 years)
 - ❑ U.S. National Science Digital Library/Distributed Learning
- ❑ **Through mid-2007, measured “use”...but it was really about “access”**
 - ❑ In the U.S. most of the projects were referatories, not repositories (contrasted with GLOBE, ARIADNE in Europe)
 - ❑ Use metrics, have been about use of metadata as pass-through to content

Log analysis

By 2007, NEEDS, SMETE.ORG, MERLOT and NSDL we were still focusing on...

▣ Simple webserver log analysis

- ▣ Number of visits to the site
- ▣ Number of visits to a metadata record/detail page
- ▣ Time spent per visit, Number of registered users (if applicable), Location of users (geographic), Time of use (day, hour), Location of referring site

So were OpenCourseWares...though MIT also had user surveys to supplement log analysis

NSDL Evaluation, Web Metrics, and Pathways Evaluator Working Groups

- ❑ **Working groups to come to common agreement and understanding on web metrics (analytics, log analysis)**
 - ❑ Very similar to work underway in OCW and OER communities
- ❑ **End result, still mostly “simplistic” usage log analysis**
 - ❑ Easy to “standardize” across multiple sites, lowest common denominator...and the bigger the number, the better, right?
 - ❑ Focused on metadata records and not the content itself

What about today?

In 2011+ Learning Analytics?

- ▣ **Focus of 2010 is still mostly web analytics**
 - ▣ Applicable to both NSDL projects and OER/OCW projects...
- ▣ **Shared web analytics (Omniture, Google Analytics)**
 - ▣ Easy to implement, and standardized
 - ▣ Potential for richer data
 - ▣ ...but are they still being used for “simplistic” web analysis
 - ▣ ...leads to big numbers...~2M visits a month (across OCW Consortium)
 - ▣ ...untapped desire to do more...

Some sites do some richer “use” analysis

- ❑ **Course Material Downloads (OpenLearn, MIT OCW)**
- ❑ **Derivative works**
 - ❑ Connexions intra-site reuse, interesting because the platform enables use/reuse within the site itself, but what about use outside the platform?
- ❑ **Are these good measures? Are they being shared?**
- ❑ **What other “rich” analysis are you doing?**
 - ❑ ...mostly we mean quantitative...

What are some interesting questions we could ask, or you are asking, that go beyond web analytics?

Scale, Quantitative
Get to “use” of the materials

What are some interesting questions we could ask, that go beyond web analytics?

- ▣ **Regarding reuse of text...**
 - ▣ How many people cut and paste text from an OCW/OER?
 - ▣ How many examples of the pasted text are available on the public web?
 - ▣ How many similar examples of the pasted text are available on the public web?
- ▣ **How does reading time of an OCW/OER web page correlate with “use”?**

What are some interesting questions we could ask, that go beyond web analytics?

▣ Or with link analysis...

- ▣ How many sites link to documents (e.g., PowerPoint slides, PDF documents, etc.) available from OCWs/OERs?
- ▣ How many OCW/OER links are shared via social bookmarking services like del.icio.us or via social networking tools like Twitter and Facebook?

*What are some examples
of how we might measure or
investigate the question of “use”?*

Tom's Flickr photos

The screenshot shows the Flickr interface for user 'caswell_tom'. The page title is 'caswell_tom's photostream / Tags / reused'. A search bar contains the text 'Search caswell_tom's photostream'. On the left, there is a link to 'See all public content tagged with reused' and a section titled 'What are tags?' explaining that tags are keywords used to categorize uploads. The main content area displays a grid of 16 photo thumbnails. The photos include: a group of people at an event, a Christmas tree, a portrait in a frame, a stadium, a modern building, a sign that says 'WELCOME DEAR FISH EATERS', a sign that says 'THROUGH YOUR DONATION PUSH THE SEASIDE BEACH FISH IN THE SEA', a red sign with a fish, a boat on water, a fountain, a fountain with a tower, a beach, and several photos of people at an event.

Reference: Caswell, T. (2009, October 20). [How I track reuse and let my flickr photos wander](http://tomcaswell.com/2009/10/20/how-i-track-reuse-and-let-my-flickr-photos-wander/). Retrieved on November 3, 2010 from Tom's Two Cents Website: <http://tomcaswell.com/2009/10/20/how-i-track-reuse-and-let-my-flickr-photos-wander/>



Scott Leslie's Work

As part of his collaboration with OLNet

▣ Usage tracking widget in web pages

- ▣ Existing production process used by faculty
- ▣ Combine as part of the licensing process, users are already asked to insert code
- ▣ Create a placeholder code, that is then replaced automatically and tracked
- ▣ Web bug

Reference: Leslie, S. (2010, July 12). [OLNet Fellowship Week 2 – Initial Thoughts on Tracking Downloaded OERs](http://www.edtechpost.ca/wordpress/2010/07/12/olnet-tracking-oer-first-stab/). Retrieved on November 3, 2010 from EdTechPost Website: <http://www.edtechpost.ca/wordpress/2010/07/12/olnet-tracking-oer-first-stab/>

Scott Leslie's Work

As part of his collaboration with OLNet

▣ Questions they can answer...

- ▣ “what are the new servers this content lives on
- ▣ how many time each page of content in the resource (depending on how extensively they have pasted the tracking code) has been viewed, both total and unique views
- ▣ other details about the end users of the content, for instance their location and other client details”

Reference: Leslie, S. (2010, July 12). [OLNet Fellowship Week 2 – Initial Thoughts on Tracking Downloaded OERs](http://www.edtechpost.ca/wordpress/2010/07/12/olnet-tracking-oer-first-stab/). Retrieved on November 3, 2010 from EdTechPost Website: <http://www.edtechpost.ca/wordpress/2010/07/12/olnet-tracking-oer-first-stab/>

CAPRETT

(Cut and Paste Reuse Tracking Tool)

- **Cut and paste from a lecture?**
 - Simple cut and paste supported from HTML pages
 - Simple cut and paste possible from PDF/Word/etc.
 - What about tynt.com-style support
 - When the user highlights text, an automatic linkback to exact location in original page is created
 - Extend tynt.com to add attribution information automatically and pasted with text

Reference: Muramatsu, B. (2009, August 27). Plagiarism is Good™ Revisited. Retrieved on May 5, 2010 from Brandon Muramatsu's Website: <http://www.mura.org/2009/08/plagiarism-is-good-revisited/>

Tynt Insight

The screenshot shows the SFGate website interface. At the top, there's a navigation bar with categories like Home, News, Sports, Business, Entertainment, Food, Living, Travel, Columns, Buy & Sell, Jobs, Real Estate, Cars, and Index. Below this is a search bar and a navigation menu for various sports teams and leagues. The main content area features a news article titled "Baylor's Udoh entering NBA draft early" dated Tuesday, April 13, 2010. The article text states: "(04-13) 16:26 PDT Waco, Texas (AP) -- Ekpe Udoh plans to bypass his senior season at Baylor and enter the NBA draft. The 6-foot-10 power forward said he hasn't decided on an agent, but does plan to hire one and be available in the June draft. Early projections have Udoh being a possible a lottery pick." To the right of the article is a "Galleries" section with three image thumbnails: "The Heavy" Premiere, "It's a Wonderful Afterlife" Premiere, and Beverly Hills Film Festival. Below the article is a "Fan Shop" section and a "MORE SPORTS" section with links to other sports news. At the bottom right, there's a promotional banner for "APRIL 23-25 vs. STL" with details for "FRI 4/23 WEARABLE BLANKET NIGHT" and "SAT 4/24 PABLO SANDOVAL BOBBLEHEAD NIGHT".

Source: tynt.com




Plagiarism Detection

- **TurnItIn**

- Checks student essays against it's database of submitted works, plus the open web

- **...but remember, we'd like to see copies of OCW/OER content...with attribution of course**

Some thoughts... Transforming Access Metrics to Use Metrics

Current		Proposed
Visitors		Take visitor counts, but pair with groupings of time/page, time/visit Visitor Categories: <ul style="list-style-type: none">• Answer a question• Learn a concept• Take a course• Just looking around
Visits or downloads		<ul style="list-style-type: none">• Appearance of text in other web-accessible documents• Web bugs to track content within downloads
None		Cut and paste trackers

The 30 second summary

- ❑ **“Plagiarism is good” is an idea we’ve been talking about for 3 years now**
- ❑ **Move beyond access metrics, to “use” metrics that many sites are capable of evaluating**
 - ❑ Access metrics are “easy”, and can be “standardized”
 - ❑ But, they’re like McDonald’s—“Billions Served”
- ❑ **So, let’s start using something else...and reporting something else...**

Perhaps you’ve heard something you want to implement in your projects...

Thank you!

Brandon Muramatsu, mura@mit.edu
Tom Caswell, [@tom4cam](mailto:caswell.tom@gmail.com)
Flora McMartin, flora.mcmartin@gmail.com