ASSOCIATION OF RESEARCH LIBRARIES®
empowering the research library community

Accuracy in Web Analytics Reporting on Digital Libraries

Patrick OBrien, Montana State University

Ricky Erway, OCLC Research

Martha Kyrillidou, Association of Research Libraries

Kenning Arlitsch, Montana State University

CNI Membership Meeting, Washington D.C.
December 9, 2013

Web analytics

Tracking, compiling, and analyzing statistics about use of a website

Source: http://is.gd/ejTqRv

Considerations

OCLC Research

- Privacy
- Usage of non-html Web assets
- Blinded by big numbers
- Spiders, robots, proxy servers, caching
- Apples to oranges
- Quantitative vs. qualitative

Why

- First the big question
- Study use of library web properties
- Site redesign planning
- Assess changes to the user interface
- Staff deployment
- Support budget needs, ROI
- User satisfaction
- Reporting to parent institution, ARL, NCES …

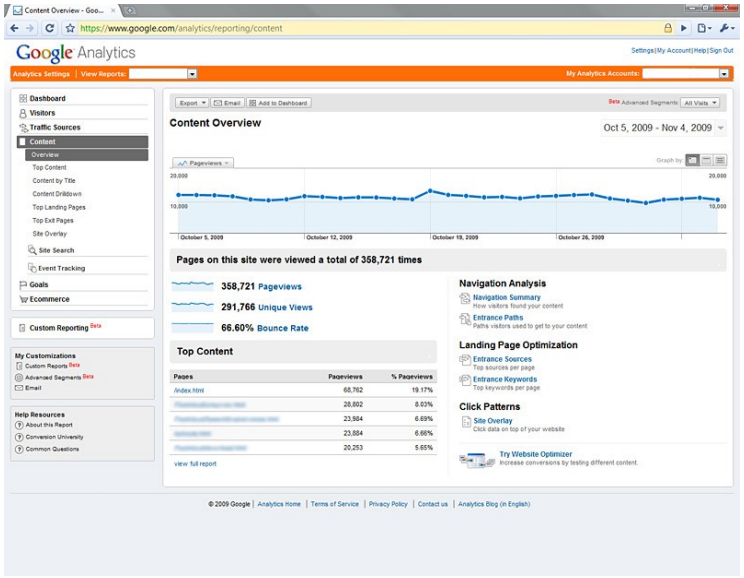
What

- What are users searching for?
- Which features are used?
- Where do our users come from?
- Which documents are downloaded?
- How long do users stay on our site?
- What are their navigation paths?
- How do we compare over time?
- How do we compare to others?

How

- Hits
- Page views
- Time on page
- Visitors (Unique? New?)
- Visits
- Return visits
- Bounce rate

Tools



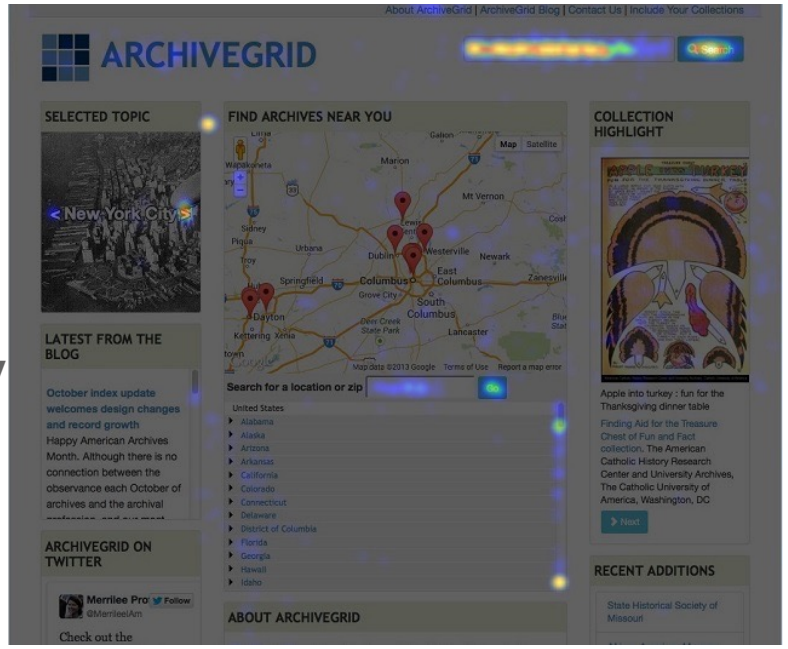
OCLC Research

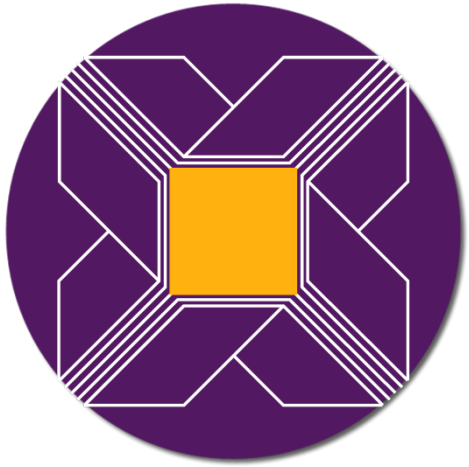
Web Analytics

- **Google Analytics**
- Adobe Analytics, IBM Digital Analytics, WebTrends …
- Piwik, AWStats… (open source)

Heatmaps

- CrazyEgg
- ClickTale, Clickdensity





ARL perspectives on accuracy in web analytics reporting on digital libraries

The need for assessment

- Underlying need to demonstrate our worth
- The reallocation of resources from traditional services and functions
- Rapid shifts in information-seeking behavior
- Increasing user demands

LibValue: an incubator
And the history of ARL tools

To describe and measure the performance of research libraries and their contribution to research, teaching, learning and community service

ARL Statistics™

Since 1907-08

LibQUAL+®

Since 2000

MINES for Libraries™

Since 2003

ClimateQUAL™

Since 2007

DigiQUAL®

ARL Statistics and web analytics

- E-metrics dating back in 2001
- Founding member of COUNTER
- Testing data elements with mixed success

- Downloads, searches, and federated searches

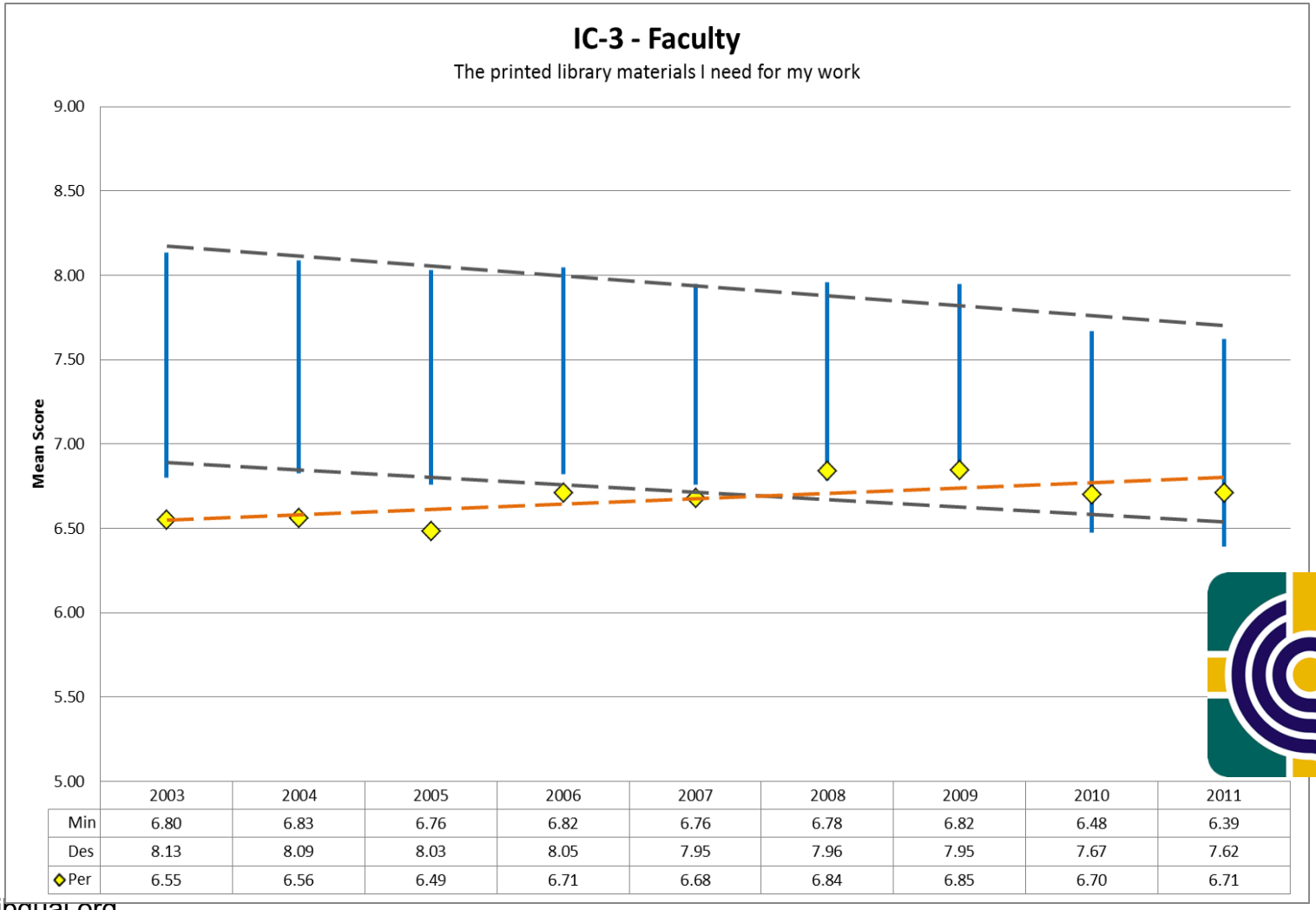
- Latest Challenge: searches from ‘discovery’ systems

… gripes … from coll-assess

- “unfortunately, because of the way Discovery systems work, they never interact directly with the source database or platform. Instead, all searches are conducted entirely within the Discovery system’s platform. Thus, there is no search to record at the source database end. So, it is not reported in the DB1 reports as Federated Searches/Sessions. This is a biiiiiiig problem for us”

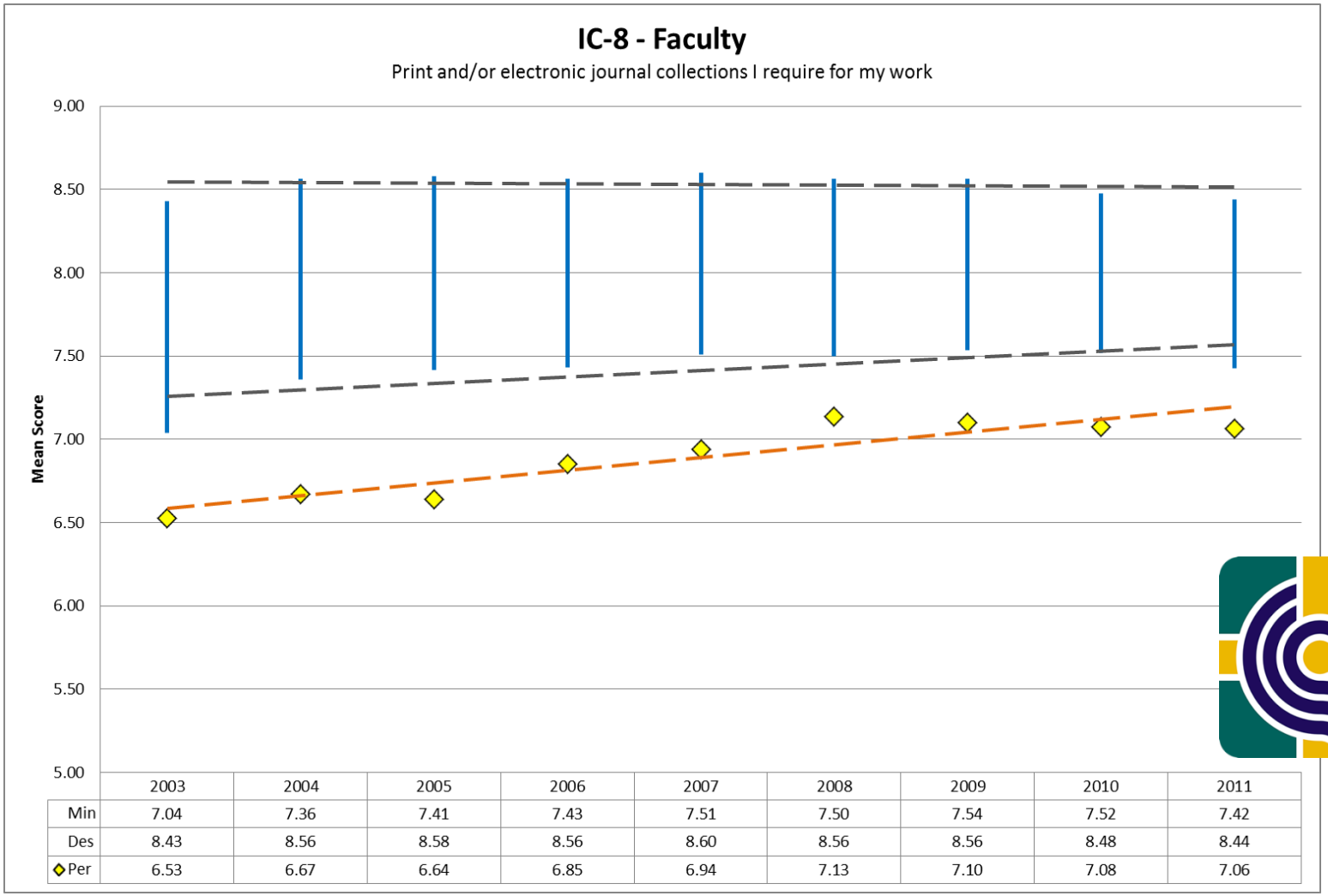
- “clients need to complain about the lack of good stats available to us and to demand something better …. I know of some institutions that have implemented Google Analytics tracking for <discovery system> so that they can collect information about the content that users click on from within the <> index. We really shouldn't have to be doing this

The Printed Library Materials I Need for My Work



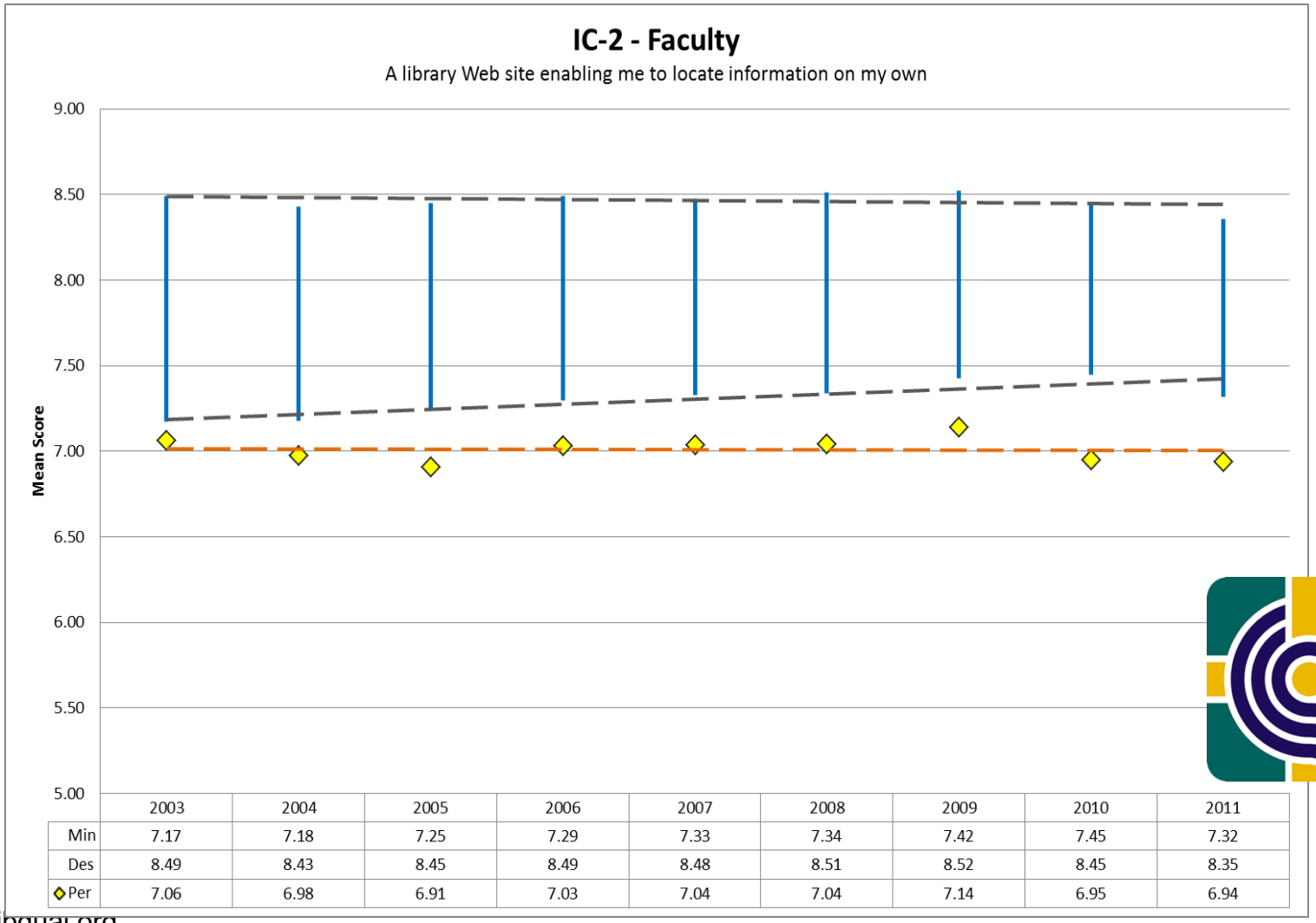
	2003	2004	2005	2006	2007	2008	2009	2010	2011
Min	6.80	6.83	6.76	6.82	6.76	6.78	6.82	6.48	6.39
Des	8.13	8.09	8.03	8.05	7.95	7.96	7.95	7.67	7.62
Per	6.55	6.56	6.49	6.71	6.68	6.84	6.85	6.70	6.71

Print and/or Electronic Journal Collections I Require for My Work



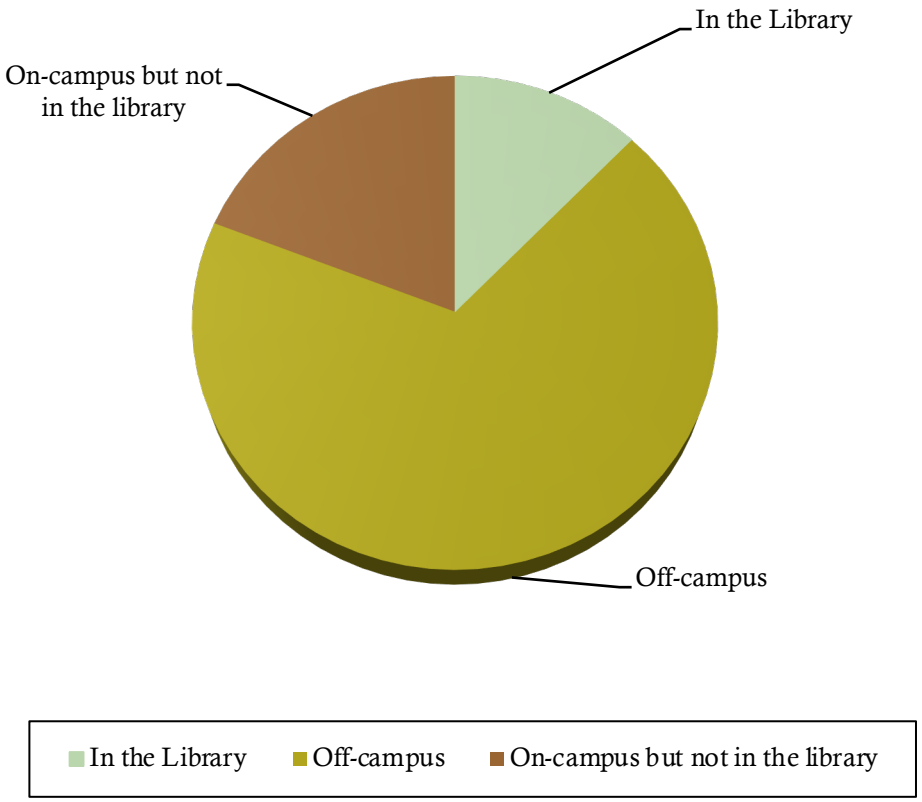
	2003	2004	2005	2006	2007	2008	2009	2010	2011
Min	7.04	7.36	7.41	7.43	7.51	7.50	7.54	7.52	7.42
Des	8.43	8.56	8.58	8.56	8.60	8.56	8.56	8.48	8.44
◇Per	6.53	6.67	6.64	6.85	6.94	7.13	7.10	7.08	7.06

A Library Website Enabling Me to Locate Information on My Own



	2003	2004	2005	2006	2007	2008	2009	2010	2011
Min	7.17	7.18	7.25	7.29	7.33	7.34	7.42	7.45	7.32
Des	8.49	8.43	8.45	8.49	8.48	8.51	8.52	8.45	8.35
Per	7.06	6.98	6.91	7.03	7.04	7.04	7.14	6.95	6.94

MINES for Libraries – Scholars Portal @ OCUL



LibValue

Special Collections – Google Analytics application

- **LibValue: Digitized Special Collections** (video on YouTube)

Thursday, August 15, 2013, 1:00–2:00 p.m. eastern

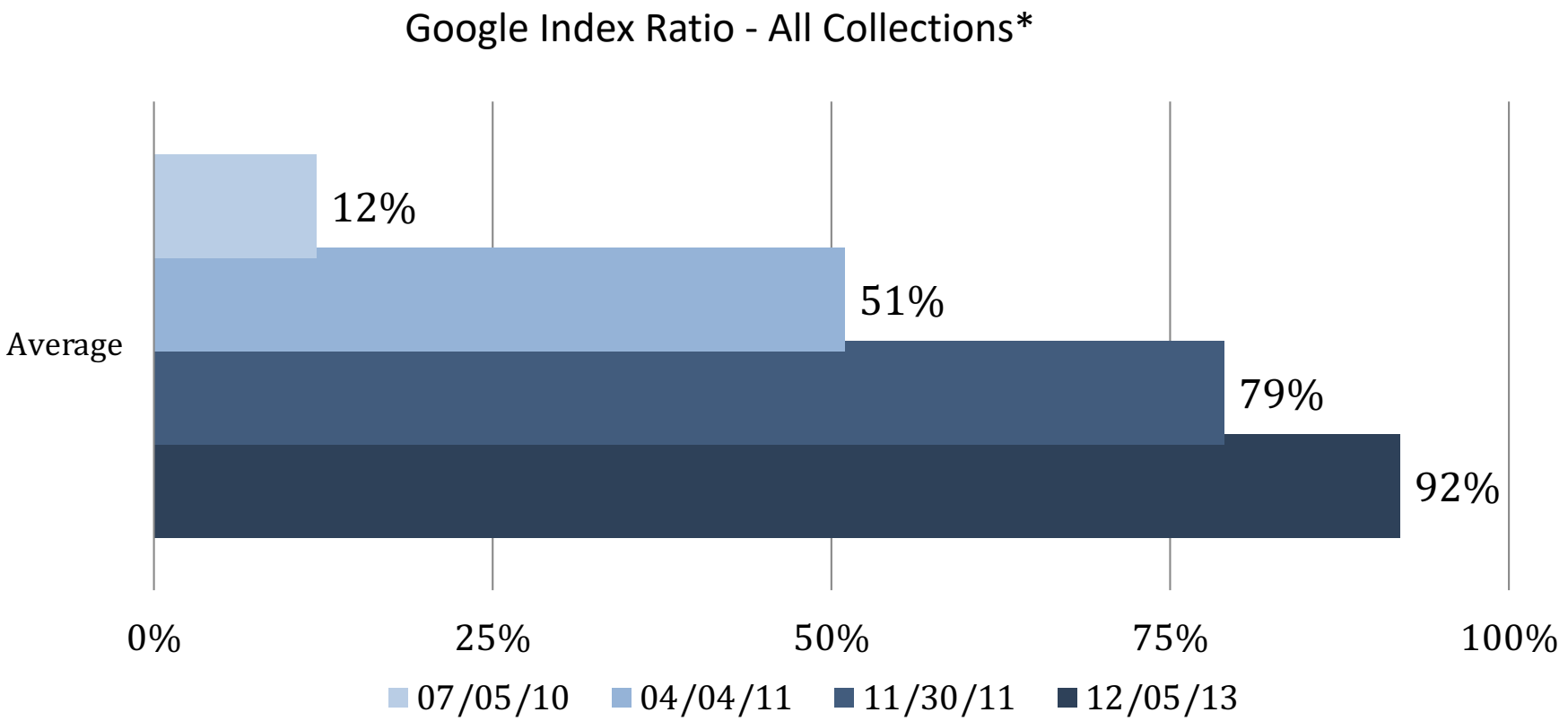
Presenters:
Gayle Baker, *Professor and Electronic Resources Coordinator, University of Tennessee Libraries*
Ken Wise, *Associate Professor, University of Tennessee Libraries*

2010: began looking for metrics on digital collection accessibility and use at Utah

- 12+ Billion
  - Number of search queries submitted to Google each month by Americans*
- 12%
  - Percentage of University of Utah digital collection content indexed by Google
- 0.5%
  - Percentage of scholarly papers in Utah's open access IR accessible to researchers via Google Scholar

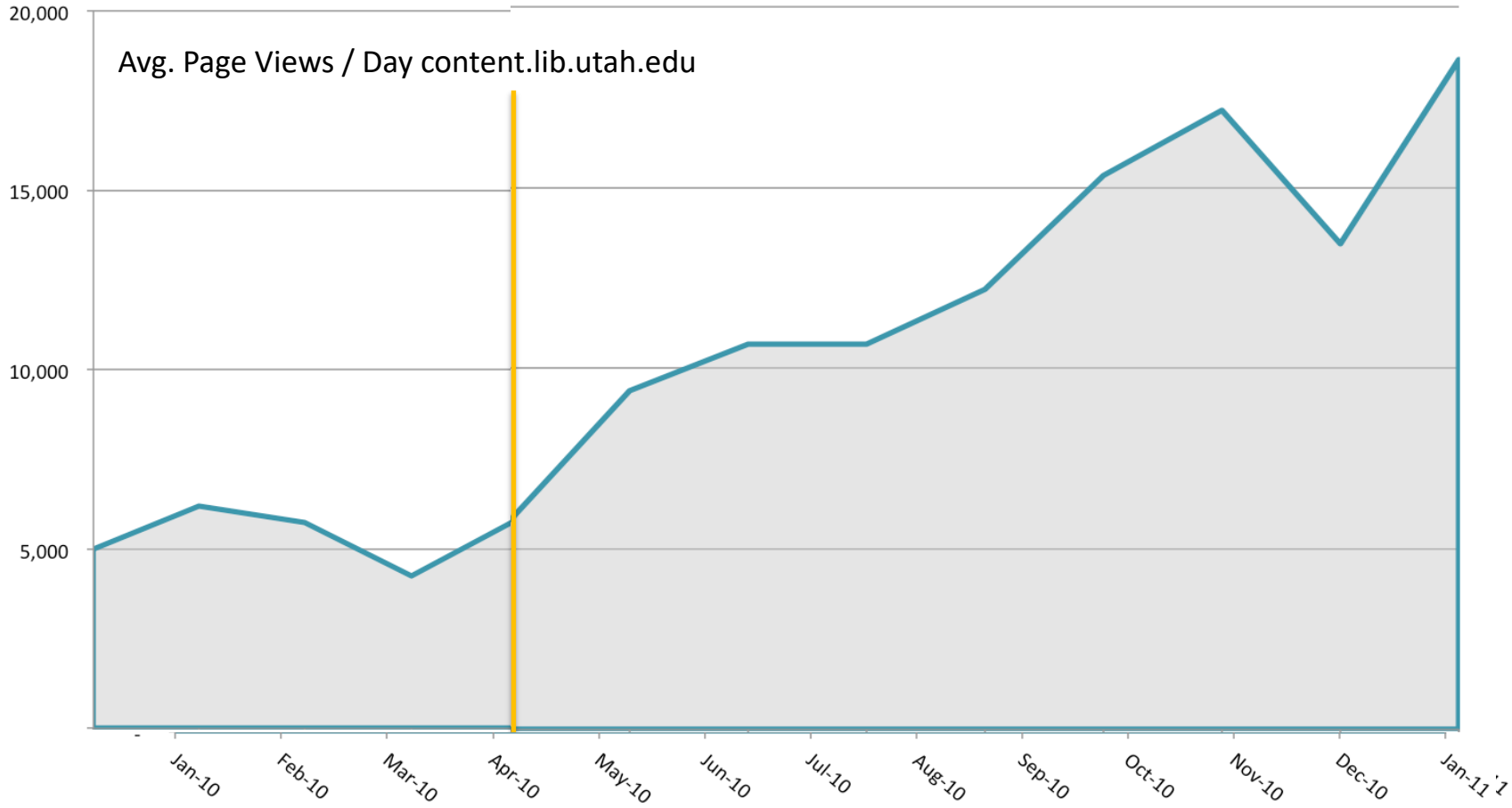
* http://www.comscore.com/Press_Events/Press_Releases/2012/1/comScore_Releases_December_2011_U.S._Search_Engine_Rankings

Basic SEO has improved collection accessibility in Google across the board…



- Google Index Ratio = URLs submitted / URLs Indexed by Google
- ~150 collections containing ~170,00 URLs (07/2010) and ~170 collections containing ~282,000 URLs (12/2013)

Producing significant increases in the average number of collection page views per day.



…resulting in more referrals and visitors

12 week comparison 2010 vs. 2012

Referring Domain: Search Results For google

Items Per Page: 10

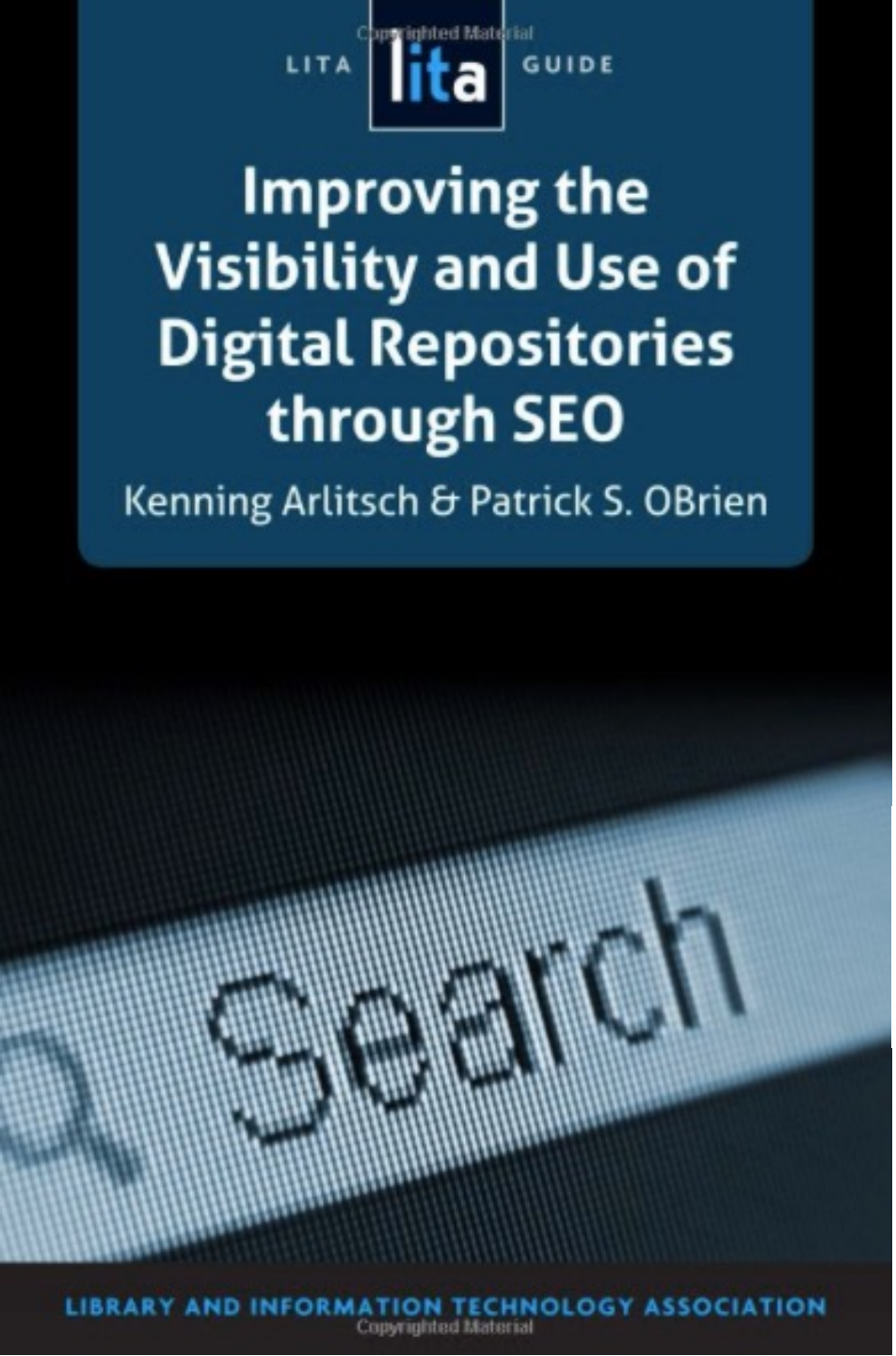
Domain	Custom View: 2/6/12 - 4/29/12 Visits	Custom View: 2/1/10 - 4/25/10 Visits	% Change Visits	% Change
1. google.com	51,694	8,959	477.01%	135.90%
2. google.co.uk	1,284	182	605.49%	145.77%
3. google.ca	1,203	415	189.88%	51.36%
4. google.it	670	38	1,663.16%	-8.87%
5. google.co.in	602	68	785.29%	
6. google.fr	475	35	1,257.14%	
7. google.es	466	26	1,692.31%	
8. google.com.au	463	95	387.37%	
9. google.de	441	88	401.14%	
10. google.com.br	408	29		
Total	**63,637**	**10,559**	Increase **502.68%**	

Themes discovered

- Traditional SEO is an afterthought
- Librarians think too small about potential traffic
- Organizational communication is poor
- Data in repositories are often messy
- Analytics are usually poorly implemented
- Vendors are slow to catch on to SEO problems
- Software tools don’t exist to implement semantic web SEO

Recommended SEO Process

1. Institutionalize SEO
   - Accurate Measurement Tools
   - Strategic Plan
2. Traditional SEO
   - Get Indexed = Index Ratio
   - Get Visible = Search Engine Results Page (SERP)
3. Semantic SEO
   - Get Relevant = Click Through Ratios (CTR)
     - Metadata
     - Linked Open Data (LOD)
     - Schema.org



Invisible institutional repositories

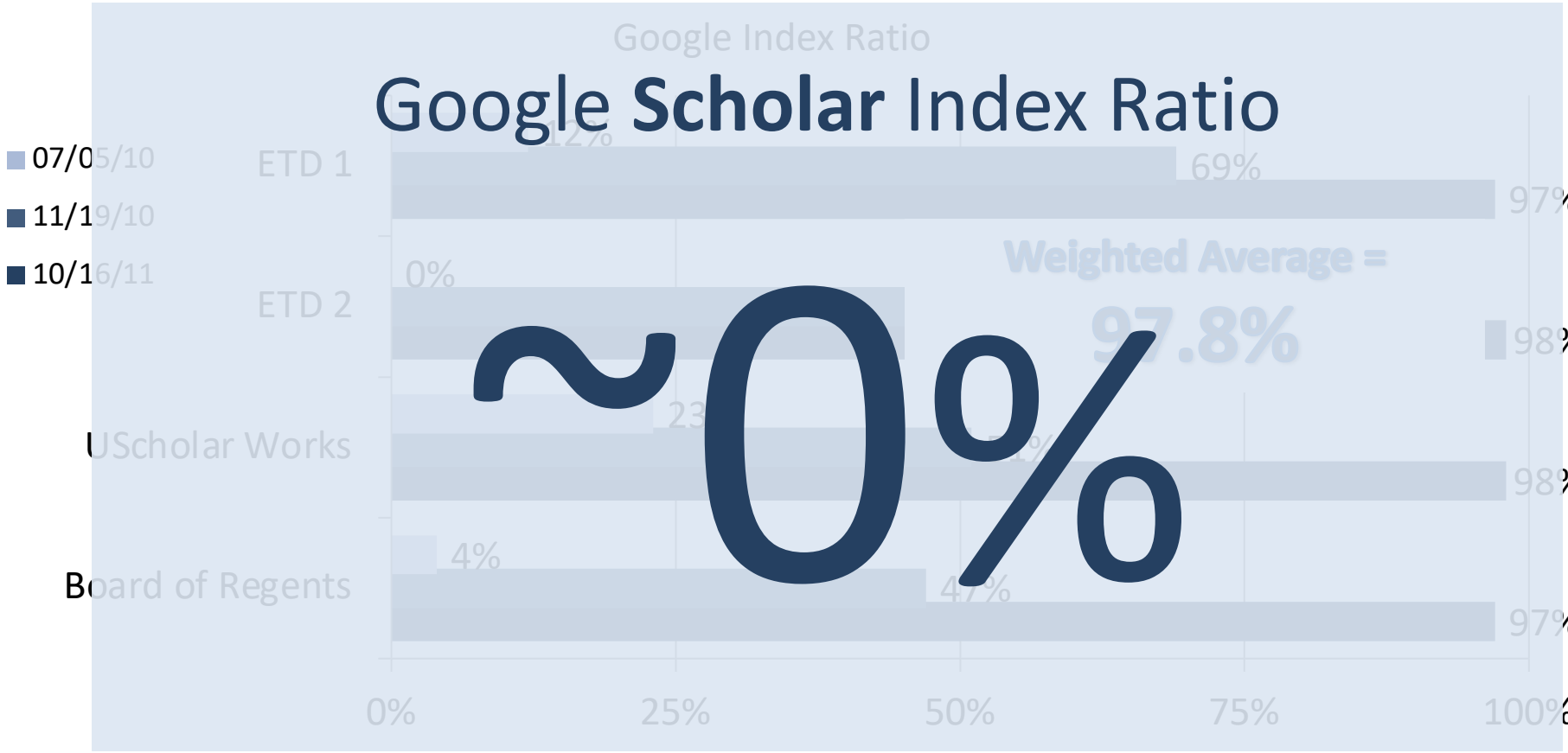
Addressing the low indexing ratios of IRs in Google Scholar

Kenning Arlitsch and Patrick S. O'Brien
J. Willard Marriott Library, University of Utah, Salt Lake City, Utah, USA

MANAGING SEARCH ENGINE OPTIMIZATION:
AN INTRODUCTION FOR LIBRARY ADMINISTRATORS

KENNING ARLITSCH, PATRICK OBRIEN, and BRIAN ROSSMANN
Montana State University Library, Bozeman, MT, USA

Google indexed ~100% of the Utah's open access IR



*October 16, 2011 Weighted Average Google Index Ratio = 97.82% (10,306/10,536).

Challenge is presenting structured data SE's can identify, parse and digest

Human Readable

Wolfinger, N. H., & McKeever, M. (2006, July). Thanks for nothing: changes in income and labor force participation for never-married mothers since 1982. In *101st American Sociological Association (ASA) Annual Meeting; 2006 Aug 11-14; Montreal, Canada* (No. 2006-07-04, pp. 1-42). Institute of Public & International Affairs (IPIA), University of Utah.

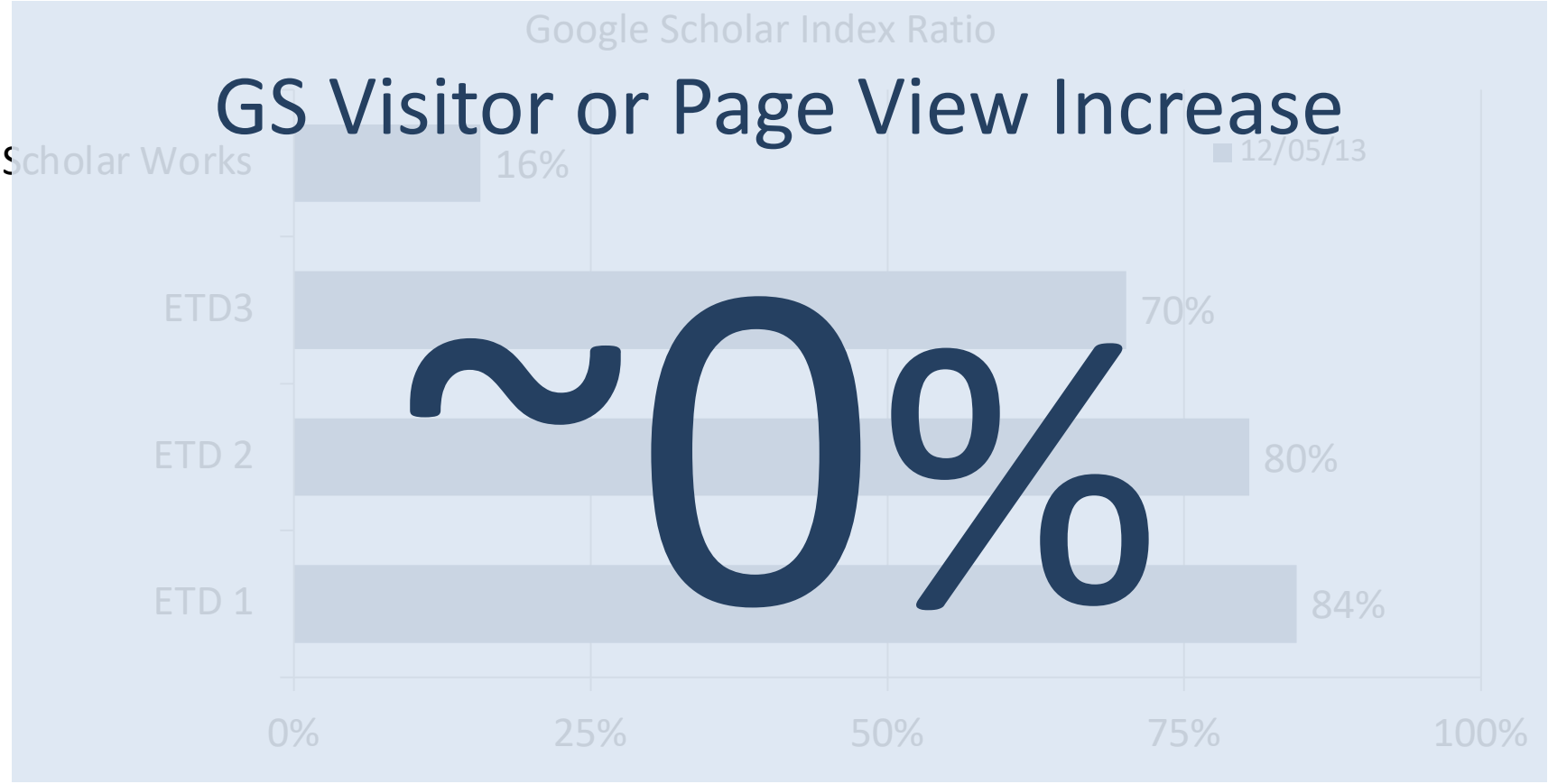
Google Scholar
Understandable

```
1  <meta name="citation_title" content="Thanks for nothing: changes in income and lab
2  <meta name="citation_author" content="Wolfinger, Nicholas H." />
3  <meta name="citation_author" content="McKeever, Matthew" />
4  <meta name="citation_date" content="2006-07-26" />
5  <meta name="citation_firstpage" content="1" />
6  <meta name="citation_lastpage" content="42" />
7  <meta name="citation_keywords" content="Motherhood; Single Mothers; Income; Popula
8  <meta name="citation_technical_report_institution" content="Institute of Public &
9  <meta name="citation_technical_report_number" content="2006-07-04" />
10 <meta name="citation_language" content="en" />
```

Google Scholar (GS) Structured Data

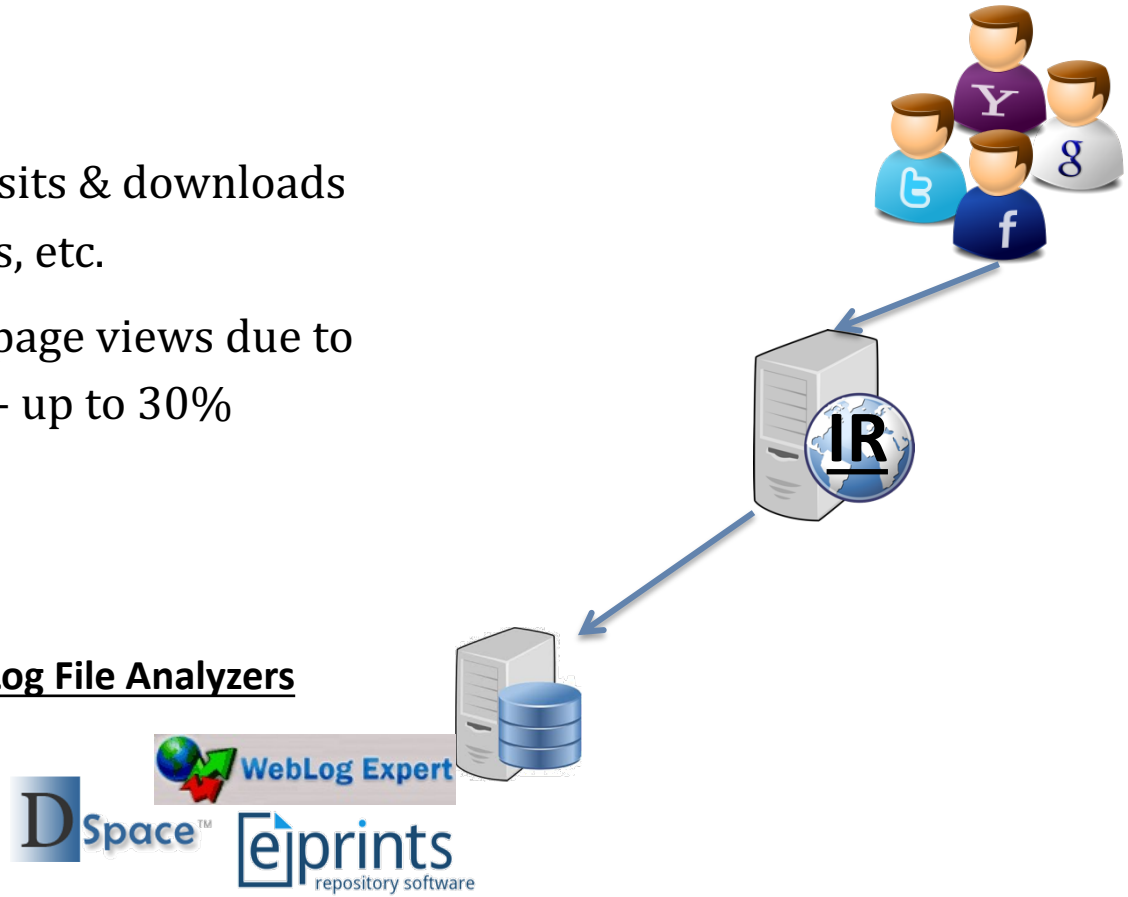
- Less than 1% of University of Utah's 8,000+ scholarly papers in GS index
- Conducted 3 pilot test
  - n=19; GS Index Ratio = 0%
  - n=19; GS Index Ratio = 62%
  - n=56; GS Index Ratio = 90%
- Metadata cleanup of 3 IR collections July 2012

Utah's open access IR items indexed by Google Scholar ~0 items to ~4,250 items



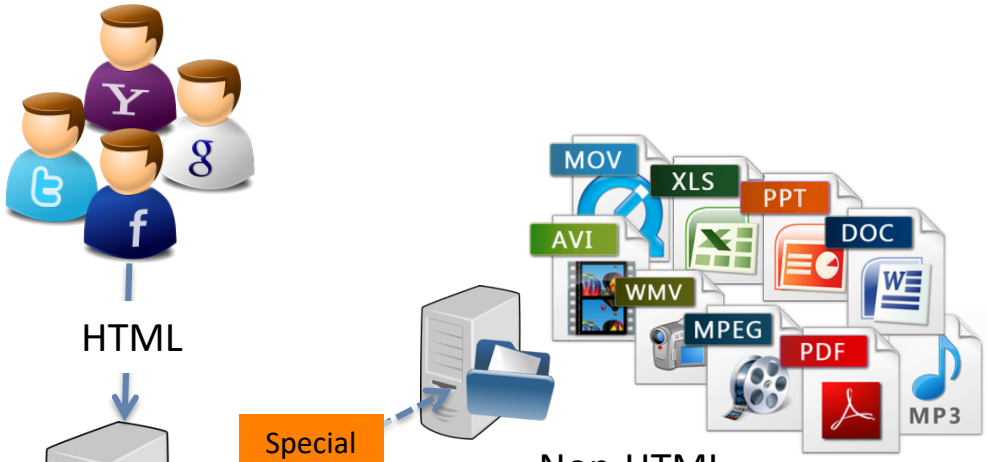
Discovered most analytics have *potential* accuracy issues for digital collections

- Log Files
  - Over count visits & downloads due to spiders, etc.
  - Under count page views due to web caching – up to 30%



Analytics Services *do not* track non-HTML downloads out of the box

- Analytics Services
  - Under count non-HTML (e.g., PDF) file downloads



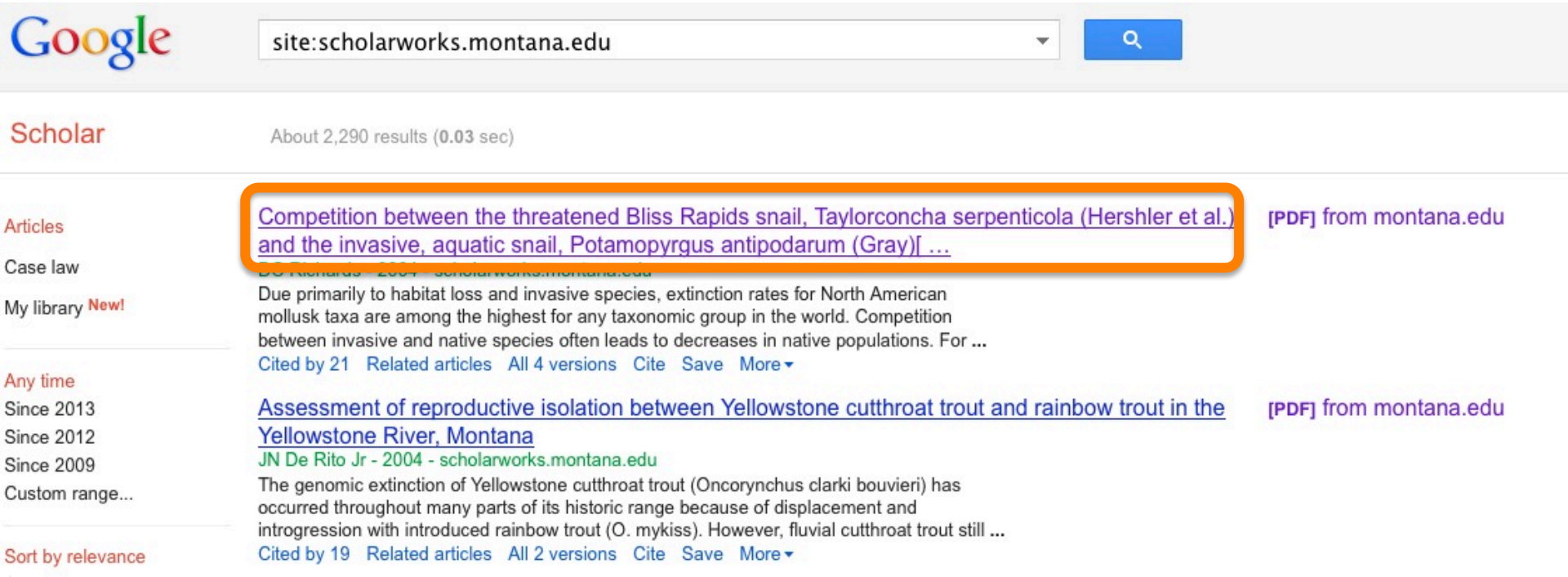
Analytics Service

OMNITURE®

Google Analytics

webtrends®

Google Scholar: HTML Request



Competition between the threatened Bliss Rapids snail, Taylorconcha serpenticola (Hershler et al.) and the invasive, aquatic snail, Potamopyrgus antipodarum (Gray) [electronic resource] / by David Charles Richards.

Richards, David Charles.

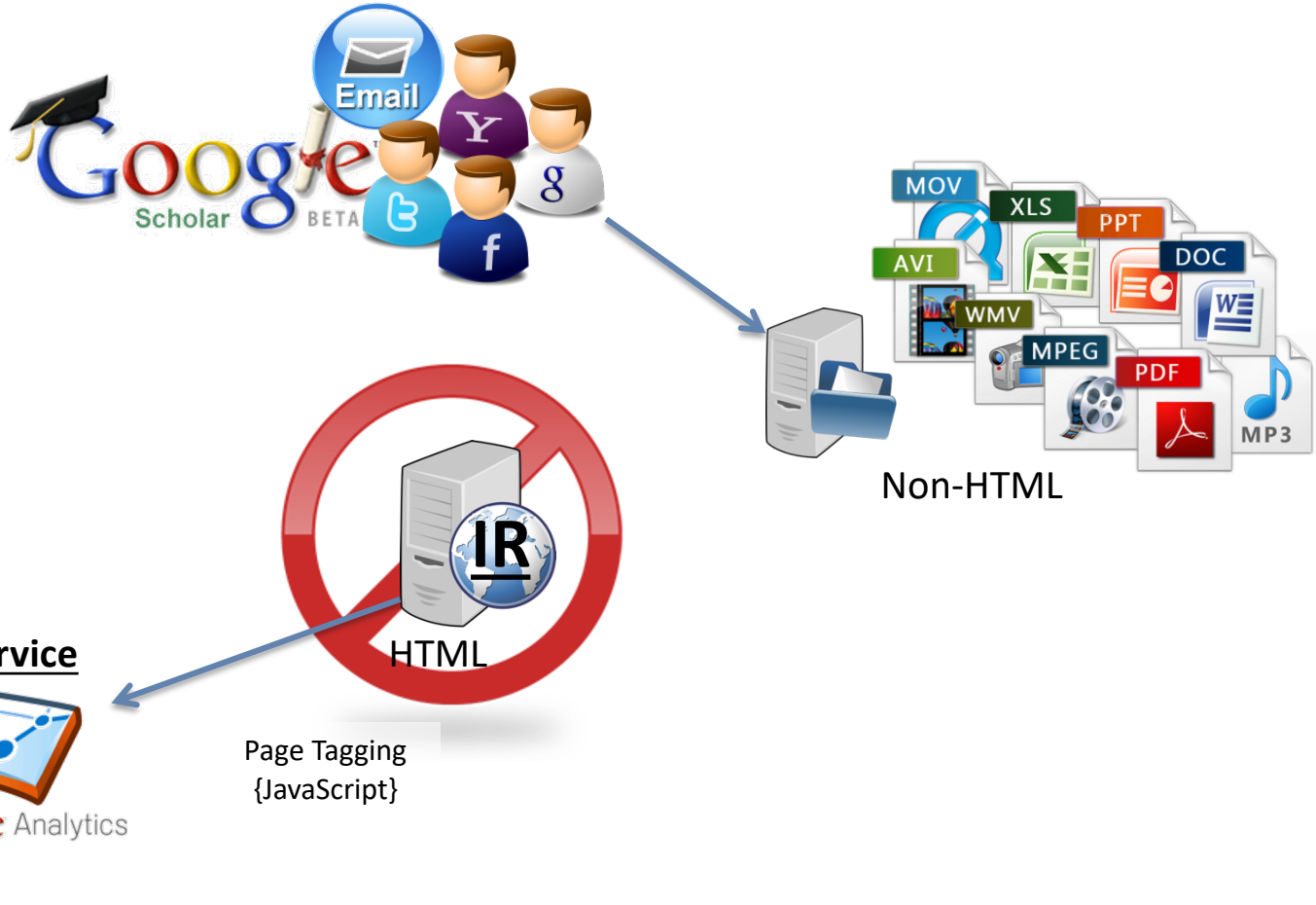
URI: http://scholarworks.montana.edu/xmlui/handle/1/2129

Date: 2004

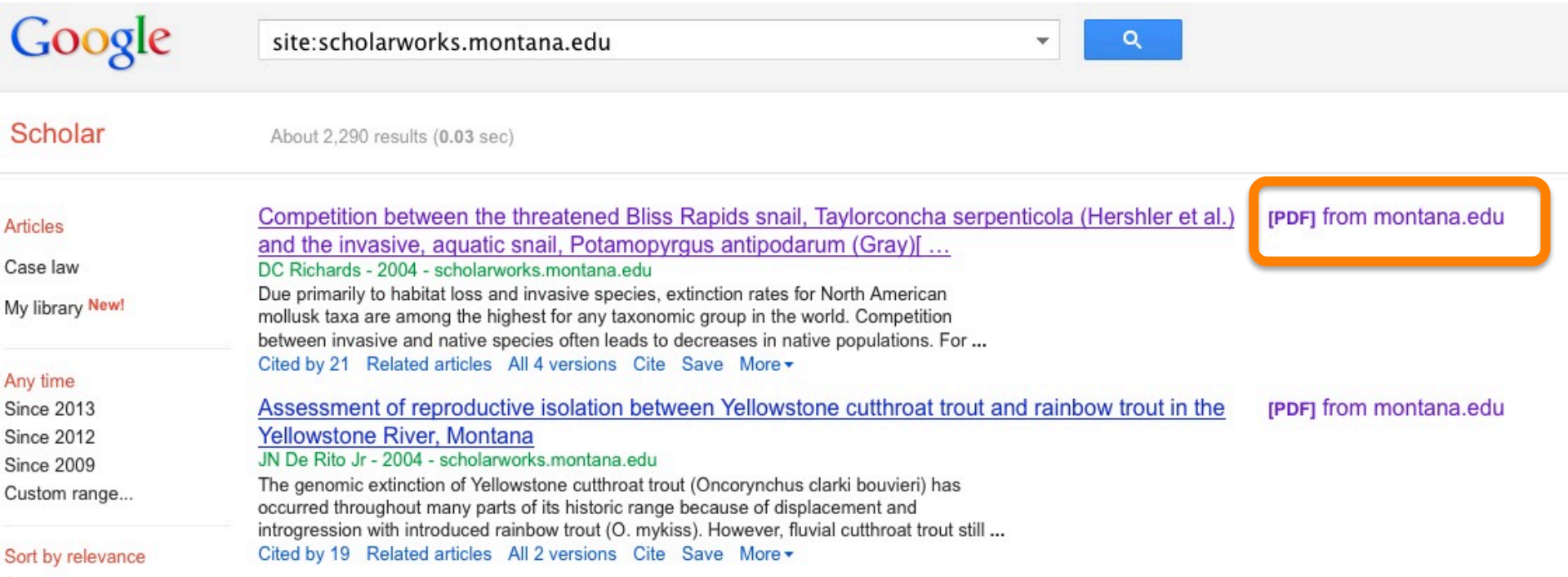
Abstract:

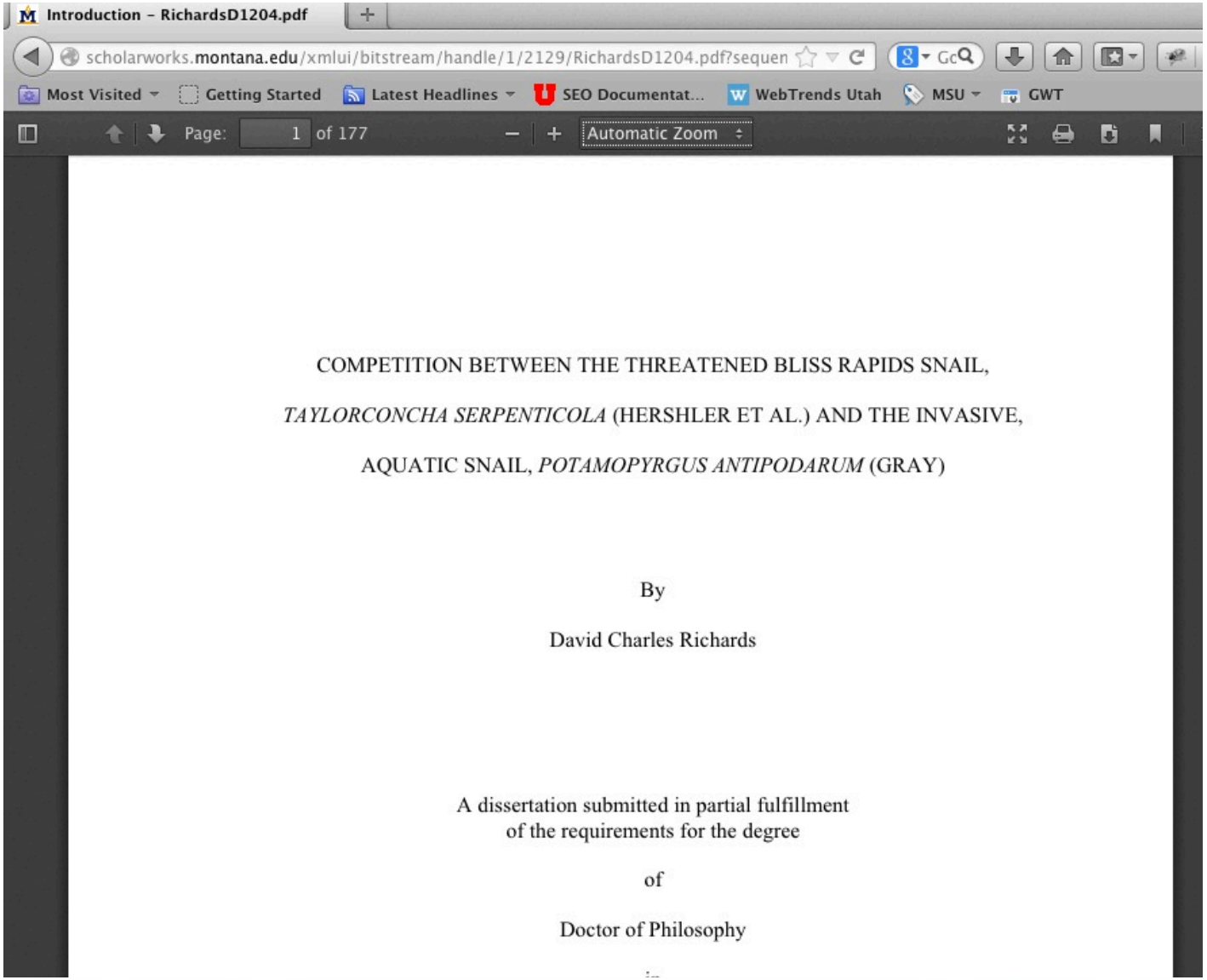
Due primarily to habitat loss and invasive species, extinction rates for North American mollusk taxa are among the highest for any taxonomic group in the world. Competition between invasive and native species often leads to decreases in native populations. For example, a primary reason for listing the Bliss Rapids snail, Taylorconcha serpenticola as threatened in the Snake River drainage was the perceived impacts of the highly invasive New Zealand mudsnail, Potamopyrgus antipodarum. Despite federal protection of T. serpenticola and the known presence of P. antipodarum in the Snake River drainage for almost 20 years, almost nothing is known about their ecology and competitive interactions. For this dissertation I conducted both field and laboratory studies to determine niche overlaps, spatial patterns, and some life history characteristics of both species. I compared optimal growth temperatures and estimated temperature tolerances for each species, under laboratory conditions; 2) examined stage (size) class fecundity rates and growth rates; and 3) examined photophobic tendencies of both species. I then explored environmental conditions and spatial patterns of both species in Banbury Springs, a tributary of the Snake River, near Hagerman, Idaho, that may have affected their distribution and abundance using regression tree analysis

Analytics Services *do not* track non-HTML file downloads via direct external links

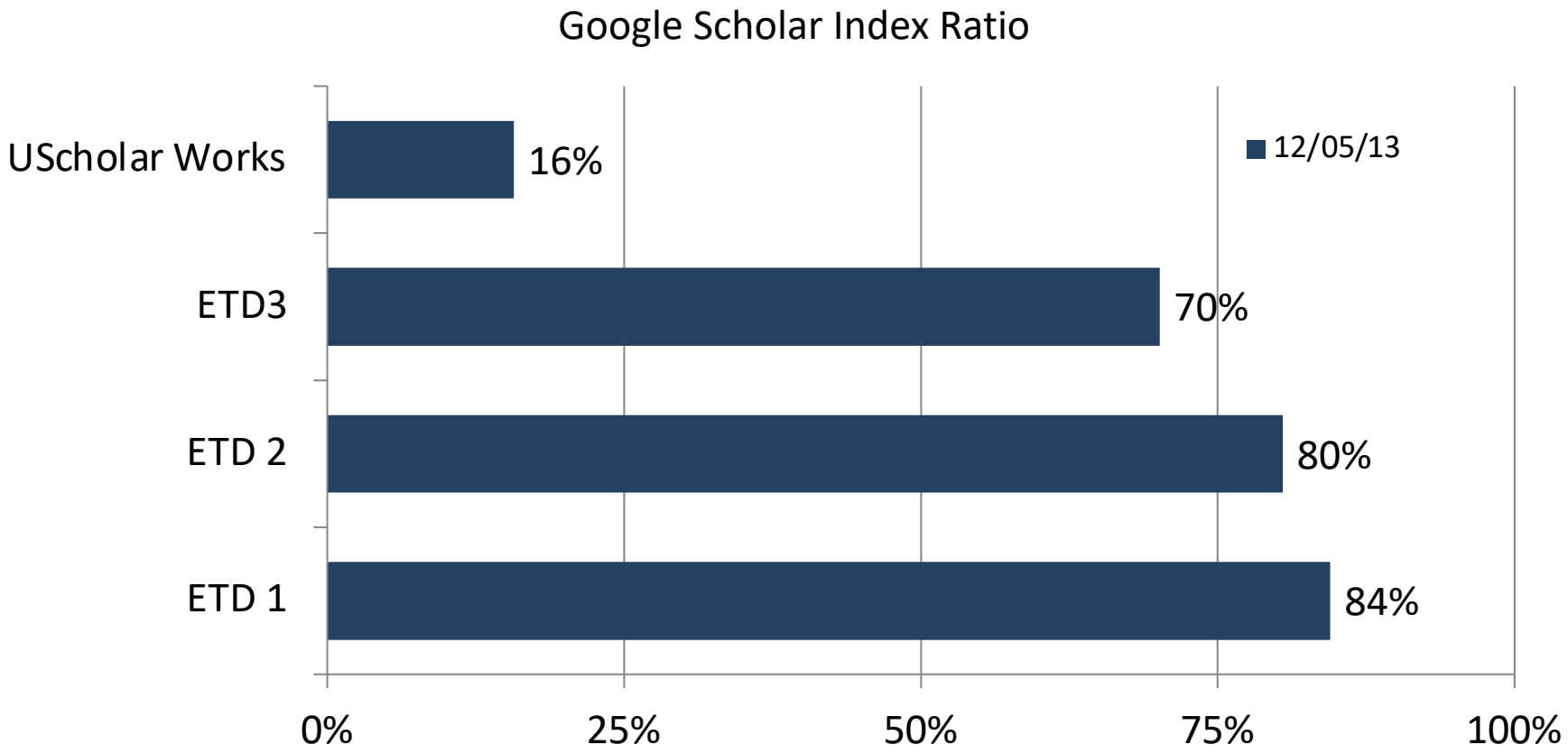


Google Scholar: PDF Request




Introduction – RichardsD1204.pdf
scholarworks.montana.edu/xmlui/bitstream/handle/1/2129/RichardsD1204.pdf?sequen
Most Visited
Getting Started
Latest Headlines
SEO Documentat...
WebTrends Utah
MSU
GWT
Page: 1 of 177
Automatic Zoom
COMPETITION BETWEEN THE THREATENED BLISS RAPIDS SNAIL,
TAYLORCONCHA SERPENTICOLA (HERSHLER ET AL.) AND THE INVASIVE,
AQUATIC SNAIL, *POTAMOPYRGUS ANTIPODARUM* (GRAY)
By
David Charles Richards
A dissertation submitted in partial fulfillment
of the requirements for the degree
of
Doctor of Philosophy

Univ of Utah open access IR items indexed by Google Scholar ~0 items to ~4,250 items



A large number Google Scholar users appear to be undercounted via Analytics services.

- 125
  - Minimum number of Google Scholar visitors invisible to Utah's open access IR
- ~200
  - Minimum number of PDF downloads by Google Scholar visitors invisible to the Utah's open access IR
- 5
  - Number of days analyzed

Montana State – ARL - OCLC partnership

- Gather more data
  - Requires additional data sets
  - Call for participation
- Develop solutions
  - Discuss policy implications
  - Training
  - Configurations

Thank You - Questions?

- Patrick OBrien – patrick.obrien4@montana.edu
- Ricky Erway – erwayr@oclc.org
- Martha Kyrillikou – martha@arl.org
- Kenning Arlitsch – kenning.arlitsch@montana.edu