




A Tale of Two^{*} Repositories

How Data Sources Affect Data Narratives


David Durden & Joseph Koivisto



*Digital Systems & Stewardship, University Libraries
University of Maryland, College Park*



Same question
different answers



What is our subject...

DRUM

Digital Repository at the University of Maryland

<https://drum.lib.umd.edu/>

What analytics platforms are we looking at...

- Apache Solr
- Google Analytics
- RAMP



```
{
  "responseHeader":{
    "status":0,
    "QTime":5287,
    "params":{
      "facet":"true",
      "indent":"true",
      "facet.query":"statistics_type:view",
      "q":"time:[2017-01-01T00:00:00Z TO 2017-01-01T24:59:59Z]",
      "wt":"json",
      "facet.pivot":"id,statistics_type"}},
  "response":{"numFound":7459,"start":0,"docs":[
    {
      "ip":"60.10.198.19",
      "dns":"60.10.198.19",
      "userAgent":"Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.2; WOW64; Trident/7.0; .NET4.0E; .NET4.0C; .NET CLR 3.5.30729; .NET CLR 2.0.50727; .NET CLR 3.0.30729)",
      "isBot":false,
      "continent":"AS",
      "countryCode":"CN",
      "city":"Hebei",
      "latitude":39.889694,
      "longitude":115.274994,
      "id":30200,
      "type":0,
      "owningItem":[8854],
      "owningColl":[119,
        1],
      "owningComm":[57,
        4,
        4,
        2],
      "time":"2017-01-01T00:01:05.421Z",
      "bundleName":["ORIGINAL"],
      "statistics_type":"view",
      "uid":"6db262b1-8a85-4055-a696-cbfd929f812b",
      "_version_":1555297065548906496},
    {
      "ip":"60.10.198.19",
      "dns":"60.10.198.19",
      "userAgent":"Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.2; WOW64; Trident/7.0; .NET4.0E; .NET4.0C; .NET CLR 3.5.30729; .NET CLR 2.0.50727; .NET CLR 3.0.30729)",
      "isBot":false,
      "continent":"AS",
      "countryCode":"CN",
      "city":"Hebei",
      "latitude":39.889694,
      "longitude":115.274994,
      "id":30200,
      "type":0,
      "owningItem":[8854],
      "owningColl":[119,
        1],
      "owningComm":[57,
        4,
        4,
        2],
      "time":"2017-01-01T00:01:06.01Z",
      "bundleName":["ORIGINAL"],
      "statistics_type":"view",
      "uid":"95f6e349-79d3-49d9-9a58-de1b5687d56b",
      "_version_":1555297066108846080},
```


When are we looking...

1 January 2017 - 30 June 2017

What data are we looking at...

Page views / Downloads





Why is this important

It depends...

- What's the question?
- What's the data?
- Basing claims on faulty, possibly over-inflated data

Understanding & Exploring Deficiencies

- What is your data source?
- What does it collect?
- In what way does it collect its data?
- What is “filtered”?



Speculations & Questions

What are some speculative approaches towards this known problem?

Learning how to properly
contextualize your data
narrative

- Be forthright about what is there & what is not
- Take the opportunity to educate stakeholders
- Disclose known problems, gaps
- Stats & analytics don't happen in a vacuum

How much should we rely on
someone else's pre-analyzed
data?

