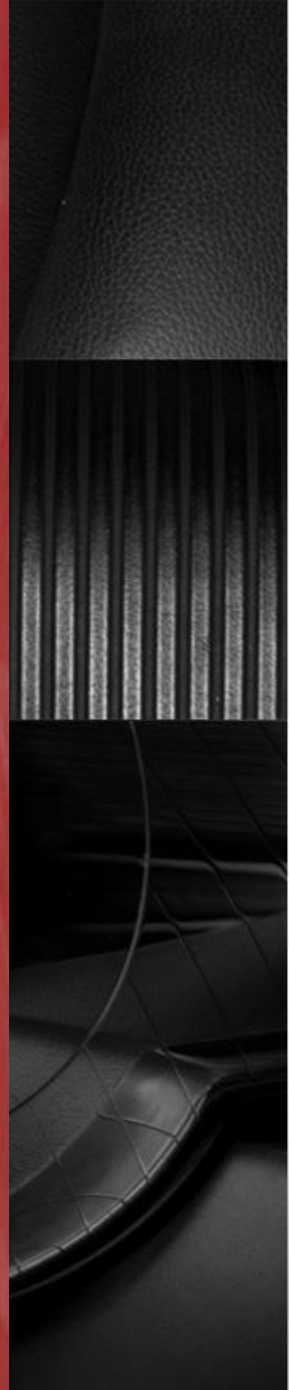


# Discovering your Discovery System in Real-Time

Annette Bailey

Godmar Back





# (Academic) Libraries in 2014: some observations

- Majority of library resources used are electronic
- Physical space is being dedicated to student work areas and digital displays
- Librarians are becoming active partners in all stages of research
- Libraries are replacing their OPACs with discovery systems



# Risks associated with these changes

- Use of library collections becomes invisible
- Search engines compete with discovery systems for users requiring substantial marketing effort for the discovery system

## Challenge:

Make the discovery process & use of resources visible!

# Can We See What Our Users are Doing?

- There's the COUNTER standard which gives us usage data
  - Usually, after the fact, and (somewhat) hard to process
  - Topic of a different talk/project ...
- Summon provides analytics data on search terms & click data
  - Aggregate, not real-time
- Google Analytics can record data
  - But no real-time access
  
- But what if we could see what our users are doing in real time (while still affording them some privacy)?
- And what if we could share that with other members of the community?



# Discovering Discovery

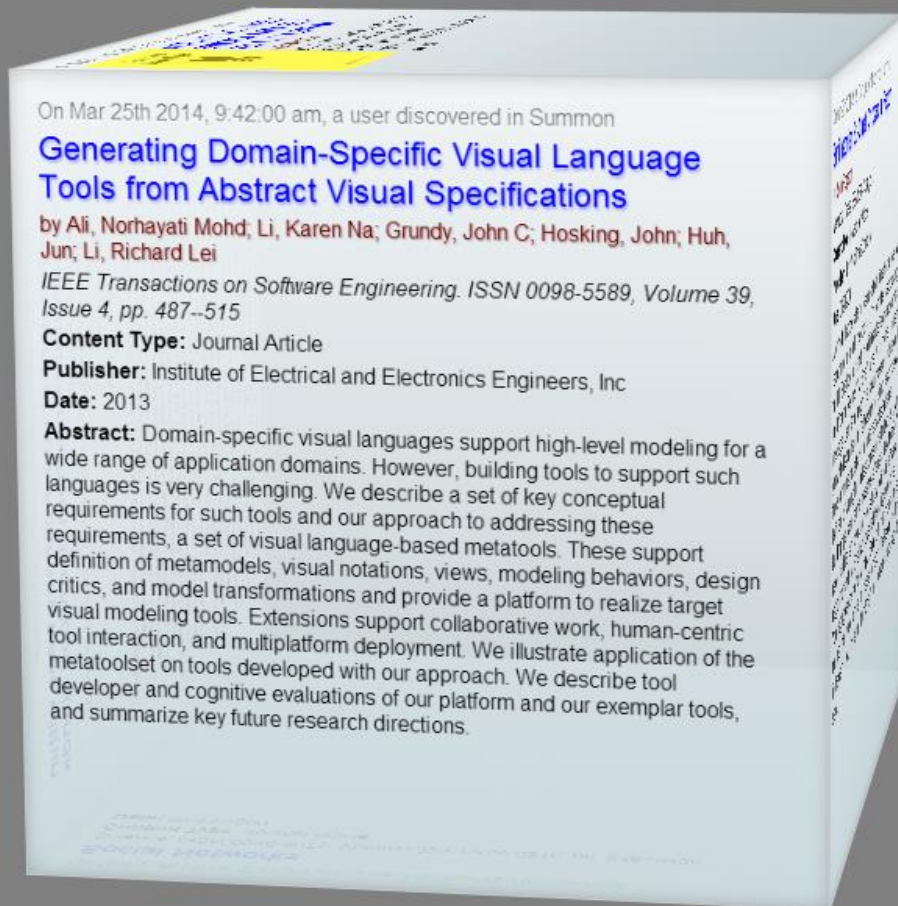
- Visualize, in real-time, the items users discover as they click on results in Summon
- It's as if we could see what object the user is plucking off the shelf. We can't see if they actually use the object, but we know that they are interested enough to go further than the search results.

# Live Demos

- Let's see what users are clicking on in the VT instance of Summon right now!
- Twitter-Style Widget: [[Link](#)]
- Summon Cube [[Link](#)]
- Word Cloud ([Subject/Last50](#)) ([Title/Last50](#))
- Gauge [[Link](#)]
- Discipline Ticker: [[Link](#)]
- Publication Year Chart [[Link](#)]



# Summon Cube

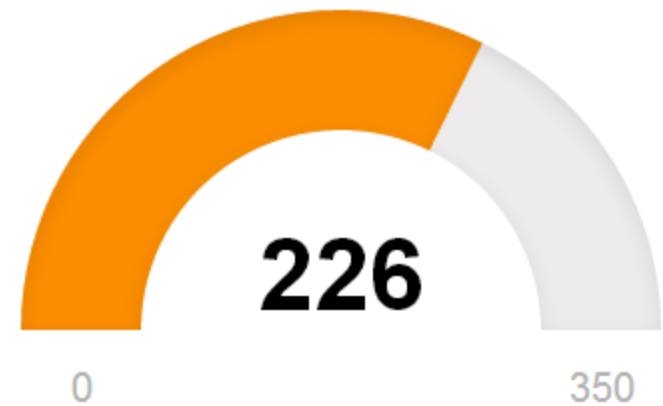




# Ticker & Gauge

## Summon Gauge

Clicks per hour



## Live ticker of Summon Disciplines

eutics, & Pharmacology: 1 (▲1) Business: 1 (▼-1) History & Archaeology: 1 (▶0) Anatomy & Physiology: 1 (▶)

This ticker shows a ticker-style histograms of the disciplines accessed in the last 50 Summon hits. The deltas shown are between the current histogram and the histogram 10 clicks ago.

# Impact

- Our research co and it is clear fro are using Summ
- This project allo members of the happening at Vi
- Intended both a





# Technical Details: Click Recording

The screenshot displays the HTML source code of a search results page. A red callout box highlights the following HTML element:

```
<div class="document" id="FETCH-vt_freecatalog_ocm088043091" type="eJournal">...</div>
```

The callout box contains the text: "Summon embeds the Summon ID of each result in the page".

Below the source code, a search result is visible for "International library of technology". The result includes the following details:

- International **library of technology**: a series of textbooks for persons engaged in the engineering professions and trades, or for those who desire information concerning them
- by International Textbook Company
- 1902
- Agriculture, Engineering, History, Technology
- eJournal: Full Text Online

The browser's developer tools show the selected element's bounding box as: `div#FETCH-vt_freecatalog_ocm088043091.document 700px x 160px`.

# Technical Details: Click Recording

```
$(function () {
  $('#resultPage .document').each(function (idx, doc) {
    var $doc = $(doc);
    var id = $doc.attr('id'); // extract id
    function recordClick() {
      var clmg = new Image(1, 1); // create unique URL to avoid browser cache
      clmg.src = "http://libx.lib.vt.edu/services/summonvis/click.gif?"
        + "id=" + encodeURIComponent(id)
        + "&_ts=" + Math.floor(Math.random() * 10000000);
    }
    // register click handler for the multiple links a user may use
    $doc.find("a.documentLink").click(recordClick);
    $doc.find("div.previewDocumentTitle a").click(recordClick);
    $doc.find("div.thumbnail a").click(recordClick);
    $doc.find("div.Availability div.summary a").click(recordClick);  });
  });
});
```

# Technical Details: Logging Clicks

- Use Apache Common Log Format (no recording of referrer URL)

```
xxx.82.xxx.75 - - [07/Nov/2013:16:20:28 -0500] "GET /services/summonvis/click.gif?id=FETCH-  
webofscience_primary_0002758417006911&_ts=390057 HTTP/1.1" 200 333  
xxx.82.xxx.92 - - [07/Nov/2013:16:21:27 -0500] "GET /services/summonvis/click.gif?id=FETCH-  
vt_catalog_b223625751&_ts=6148119 HTTP/1.1" 200 333  
xxx.82.xxx.11 - - [07/Nov/2013:16:21:27 -0500] "GET /services/summonvis/click.gif?id=FETCH-  
proquest_abstracts_14361167591&_ts=1597476 HTTP/1.1" 200 333
```

```
# Excerpt from Apache config file: /etc/apache2/sites-enabled/000-default
```

```
# do not log user's query terms
```

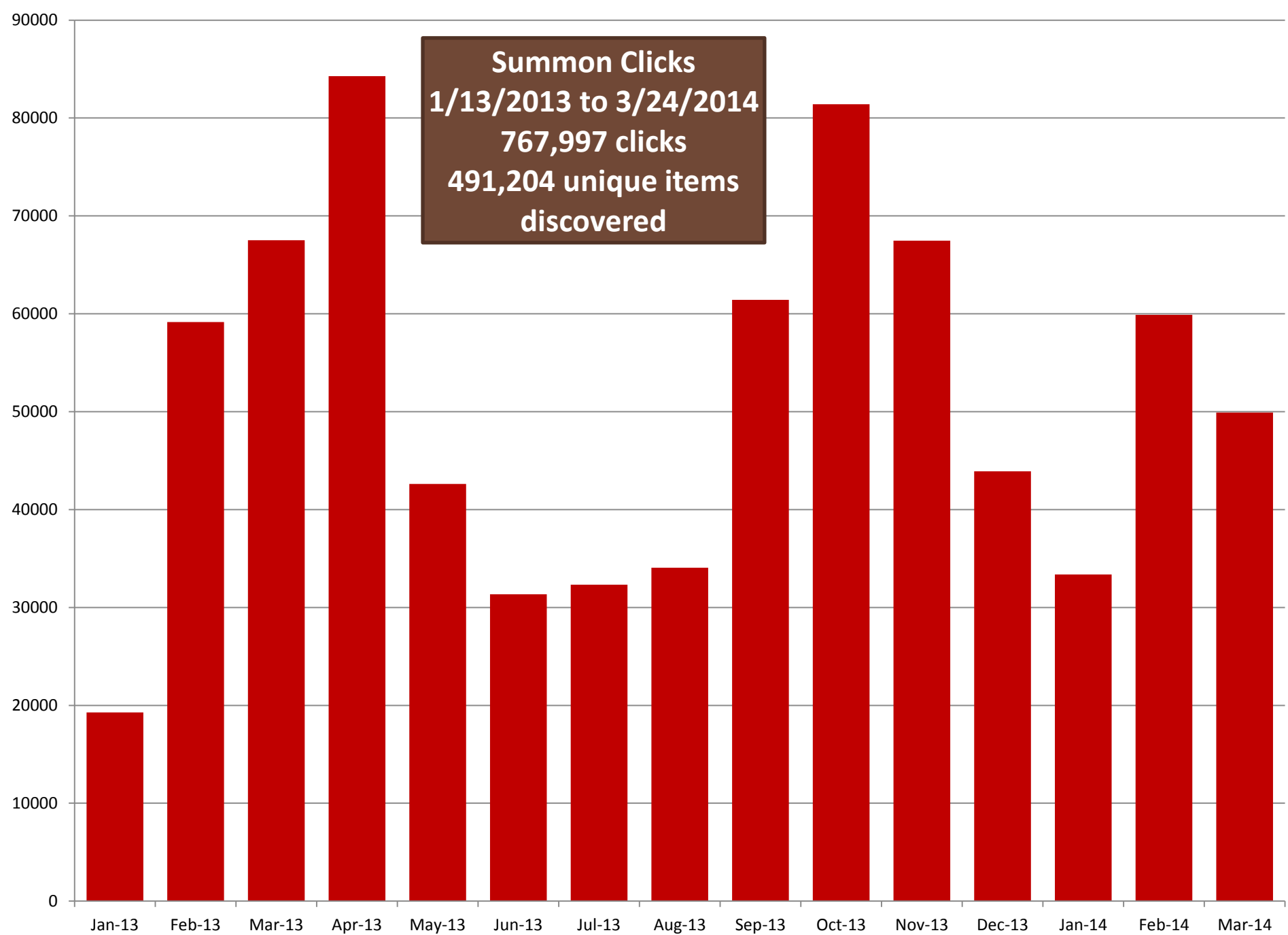
```
SetEnvIf Request_URI "^/services/summonvis" dontlog
```

```
# log resource ids users actually click on
```

```
SetEnvIf Request_URI "/services/summonvis/click.gif" clicklog
```

```
CustomLog /var/log/apache2/access.log combined env=!dontlog
```

```
CustomLog /var/log/apache2/click.log common env=clicklog
```



# Log Tracing

- When a click entry is added, we must contact Summon API right away
  - Summon IDs are very short-lived due to constant record update & merging
- Summon API
  - Supports querying in the same manner as Summon web front-end
  - Supports additional query-styles, such as retrieve by ID
  - Requires API-Key (which must be safe-guarded)
  - Returns JSON or XML
- SQLite
  - We store all records in SQLite DB
  - See <http://www.sqlite.org/>

```
import sqlite3
conn = sqlite3.connect(dbfilename)
conn.execute("""CREATE TABLE IF NOT EXISTS summonrecords
              (id TEXT PRIMARY KEY, recordjson TEXT)
              """)
```



# Sample Record

- Selected fields from JSON response



<http://en.wikipedia.org/wiki/Tanager>

```
{
  "Author": [
    "Zdziarski, JM",
    "Little, SE",
    "Adkesson, MJ"
  ],
  "ContentType": [
    "Journal Article"
  ],
  "DOI": [
    "10.1605/01.301-0000384790.2006"
  ],
  "DatabaseTitle": [
    "ProQuest Illustrata: Natural Sciences",
    "ProQuest Deep Indexing: Biological Science"
  ],
  "Discipline": [
    "Veterinary Medicine"
  ],
  "EISSN": [
    "1937-2825"
  ],
  "SubjectTerms": [
    "Cyanerpes caeruleus",
    "Cyanerpes cyaneus",
    "Tanager whole blood",
    "Polymerase chain reaction results",
    "Dacnis cayana",
    "Atoxoplasmosis In Tanagers"
  ],
  "Title": [
    "Atoxoplasmosis In Tanagers"
  ],
}
```

```
"PublicationYear": [
  "2005"
],
"SourceType": [
  "Index Database"
],
"StartPage": [
  "265"
],
"EndPoint": [
  "272"
],
"ISSN": [
  "1042-7260"
],
"IsPeerReviewed": [
  "true"
],
"IsScholarly": [
  "true"
],
"Issue": [
  "2"
],
"Language": [
  "English"
],
...
}
```

## Sample Record (2)

- Selected fields from JSON response

```
{
  "Abstract": [
    "Introduction. To report a live birth following egg retrieval after only 12 hours from hCG priming. Patients. A childless couple with five-years-lasting secondary infertility. Methods. IVF was performed according to the long protocol. Two immature oocytes were retrieved following only 12 hours after hCG priming due to the patient misunderstanding. The eggs were cultured in vitro and ICSI was performed following polar body extruded after 24 hours in culture. After additional 24 hours a 4-cell embryo was developed and ET was performed. Results. A viable pregnancy was achieved and a healthy baby girl was delivered at 38 weeks of gestation. Conclusion. In a rare and unexpected situation when immature oocytes are retrieved following a short hCG priming, the eggs should be cultured in vitro, late ICSI should be performed, and a pregnancy may be expected."
  ],
  "Snippet": [
    "\u00a0 Introduction. To report a live birth following egg retrieval after only 12 hours from hCG priming. Patients. A childless couple with five-years-lasting...",
    "Introduction. To report a live birth following egg retrieval after only 12 hours from hCG priming. Patients. A childless couple with five-years-lasting..."
  ],
  "Title": [
    "A Live Birth Subsequent to IVF following Egg Retrieval Only 12 Hours after hCG Priming"
  ],
  "URI": [
    "http://dx.doi.org/10.1155/2013/634385",
    "http://search.proquest.com/docview/1428017900",
    "http://www.ncbi.nlm.nih.gov/pubmed/23762684"
  ],
}
```

# Current Log Analysis

- Extracts & tabulates easy-to-analyze variables
  - over time periods: 1 min, 5 min, 1 hours, 1 day, 1 week
  - over counts: last 50, last 100, last 200, last 1000, last 10000
- Tabulates frequency by categories such as
  - Discipline
  - ContentType
  - SourceType (of record)
  - PublicationYear
- Tabulates word frequencies for
  - Abstract, Title, Abstract & Title together (individual words)
  - Keywords, Subject Terms, Keywords & Subject Terms together (preserving compounds)

# Example 1: Tabulating Disciplines

- Last 1 hour

```
{ Discipline:  
  { Psychology: 4,  
    'Applied Sciences': 1,  
    'Public Health': 1,  
    Forestry: 1,  
    Religion: 6,  
    'Environmental  
      Sciences': 4,  
    Music: 2,  
    Medicine: 7,  
    Mathematics: 2,  
    'Sciences (General)': 2,  
    Biology: 5,  
    Business: 1,  
    Government: 2,  
    Philosophy: 6,  
    Anthropology: 1,  
    'Computer Science': 3,
```

```
'Sociology & Social History': 2,  
  Law: 5,  
  'Languages & Literatures': 6,  
  'International Relations': 1,  
  'Library & Information Science': 1,  
  Economics: 4,  
  Physics: 4,  
  'Visual Arts': 1,  
  Ecology: 3,  
  'History & Archaeology': 1,  
  Geology: 2,  
  'Anatomy & Physiology': 1,  
  Engineering: 7,  
  Zoology: 1,  
  Education: 5,  
  Agriculture: 6,  
  Film: 1,  
  Geography: 1 }  
timestamp: '2013-11-07T19:07:00-  
05:00' }
```

# Example 2: Tabulating PubYear

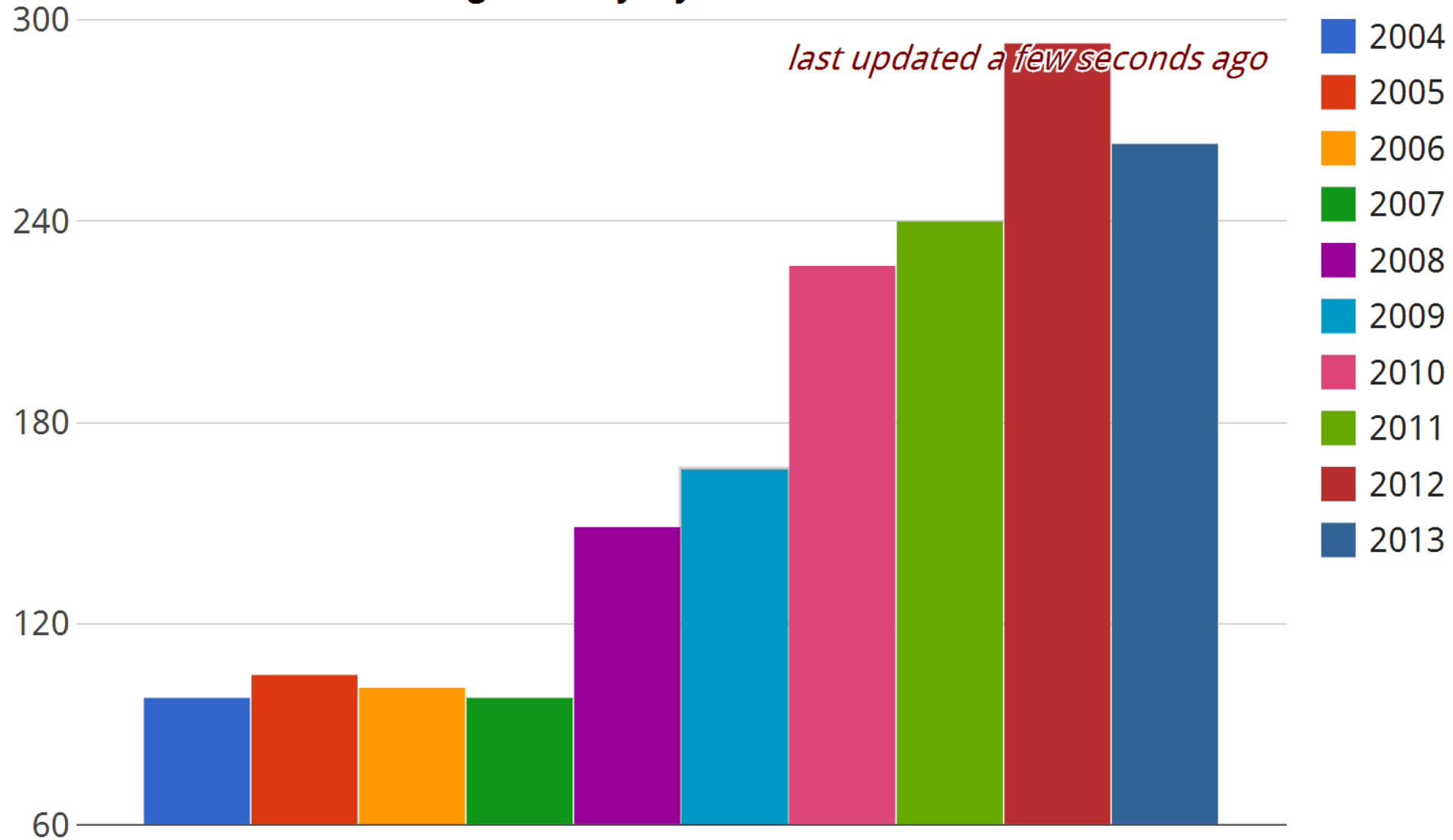
- Last 1 Day

```
{ timestamp: '2013-11-07T19:11:24-05:00',  
  PublicationYear:  
  { '1765': 1,  
    '1775': 1,  
    '1790': 1,  
    '1791': 1,  
    '1793': 2,  
    '1850': 1,  
    '1855': 1,  
    '1863': 1,  
    '1871': 1,  
    '1876': 1,  
    '1879': 1,  
    '1880': 1,  
    '1881': 1,  
    '1882': 1,  
    '1889': 1,  
    '1895': 2,  
    '1899': 3,  
    '1900': 1,  
    '1901': 4,  
    '1902': 1,  
    '1903': 1,  
    '1905': 1,  
    '1913': 1,  
    '1915': 1,  
    '1916': 2,  
    '1920': 3,  
    '1923': 1,  
    '1924': 1,  
    '1925': 1,  
    '1926': 2,  
    '1930': 2,  
    '1931': 2,  
    '1932': 2,  
    '1933': 1,  
    '1934': 1,  
    '1936': 3,  
    '1938': 1,  
    '1941': 2,  
    '1942': 2,  
    '1946': 1,  
    '1947': 1,  
    '1948': 2,  
    '1951': 1,  
    '1952': 1,  
    '1954': 3,  
    '1955': 3,  
    '1956': 2,  
    '1957': 4,  
    '1958': 7,  
    '1959': 4,
```

```
'1960': 3,  
'1961': 9,  
'1962': 5,  
'1963': 1,  
'1964': 5,  
'1965': 9,  
'1966': 9,  
'1967': 8,  
'1968': 3,  
'1969': 13,  
'1970': 10,  
'1971': 5,  
'1972': 8,  
'1973': 2,  
'1974': 16,  
'1975': 6,  
'1976': 5,  
'1977': 8,  
'1978': 4,  
'1979': 8,  
'1980': 9,  
'1981': 7,  
'1982': 10,  
'1983': 12,  
'1984': 10,  
'1985': 16,  
'1985': 16,  
'1986': 19,  
'1987': 16,  
'1988': 12,  
'1989': 21,
```

```
'1990': 29,  
'1991': 16,  
'1992': 32,  
'1993': 29,  
'1994': 29,  
'1995': 34,  
'1996': 43,  
'1997': 33,  
'1998': 62,  
'1999': 58,  
'2000': 56,  
'2001': 85,  
'2002': 72,  
'2003': 70,  
'2004': 106,  
'2005': 100,  
'2006': 104,  
'2007': 103,  
'2008': 149,  
'2009': 177,  
'2010': 222,  
'2011': 251,  
'2012': 323,  
'2013': 281,  
'2014': 1 } }
```

## Discovered Items During Last Day, By Publication Year



Generated using Google Charts [[Link](#) to live version]

# Example 3: Tabulating Title

- Last 1 Hour
- Single words
- Lowercase

```
{ timestamp: '2013-11-07T20:47:26-05:00',  
  Title:  
    [ ['special', 14 ],  
      [ 'olympics', 13 ],  
      [ 'fundraising', 10 ],  
      [ 'cannabis', 10 ],  
      [ 'food', 9 ],  
      [ 'united', 7 ],  
      [ 'history', 7 ],  
      [ 'intake', 6 ],  
      [ 'sweeteners', 6 ],  
      [ 'genetically', 6 ],  
      [ 'sex', 6 ],  
      [ 'energy', 6 ],  
      [ 'social', 6 ],  
      [ 'transsexuality', 6 ],  
      [ 'children', 6 ],  
      [ 'changed', 6 ],  
      [ 'modified', 6 ],  
      [ 'effects', 5 ],  
      [ 'states', 5 ],  
      [ 'discourse', 5 ],  
      [ 'meyerowitz', 5 ],  
      [ 'trafficking', 5 ],
```

```
    [ 'north', 4 ],  
    [ 'mentoring', 4 ],  
    [ 'medical', 4 ],  
    [ 'narrative', 4 ],  
    [ 'cbd', 4 ],  
    [ 'joanne', 4 ],  
    [ 'journal', 4 ],  
    [ 'state', 4 ],  
    [ 'mexico', 4 ],  
    [ 'era', 3 ],  
    [ 'america', 3 ],  
    [ 'economic', 3 ],  
    [ 'education', 3 ],  
    [ 'artificial', 3 ],  
    [ 'peace', 3 ],  
    [ 'high', 3 ],  
    [ 'security', 3 ],  
    [ 'treatment', 3 ],  
    [ 'weight', 3 ],  
    [ 'bugs', 3 ],  
    [ 'stink', 3 ],  
    [ 'cancer', 3 ],  
    [ 'students', 3 ],  
    [ 'theory', 3 ],  
    ...  
  }
```





# Visualization Implementation

# HTML

```
<svg xmlns="http://www.w3.org/2000/svg" viewBox="0 0 512 512">  
<title>HTML5 Logo</title>
```

```
<polygon fill="#E44D26" points="107.644,470.877 74.633,100.62 437.367,100.62
```

```
<polygon fill="#F16529" points="256,480.523 376.03,447.246 404.27,130.894 25
```

```
<polygon fill="#EBEBEB" points="256,268.217 195.91,268.217 191.76,221.716 25  
255.843,176.305 142.132,176.305 143.219,188.488 154.38,313.627 256,313.627"
```

```
<polygon fill="#EBEBEB" points="256,386.153 255.801,386.206 205.227,372.55 2  
156.409,336.333 162.771,407.634 255.791,433.457 256,433.399"/>
```

```
<path d="M108.382,0h23.077v22.8h21.11V0h23.078v69.044H152.57v-23.12h-2:
```

```
<path d="M205.994,22.896h-20.316V0h63.72v22.896h-20.325v46.148h-23.078V
```

```
<path d="M259.511,0h24.063l14.802,24.26L313.163,0h24.072v69.044h-22.982v  
15.888-24.549v34.222h-22.58V0z"/>
```

```
<path d="M348.72,0h23.084v46.222h32.453v22.822H348.72V0z"/>
```

```
<polygon fill="FFFFFF" points="255.843,268.217 255.843,313.627 311.761,313.627 306.49,372.521 255.843,386.191  
255.843,433.435 348.937,407.634 349.62,399.962 360.291,280.411 361.399,268.217 349.162,268.217"/>
```

```
<polygon fill="FFFFFF" points="255.843,176.305 255.843,204.509 255.843,221.605 255.843,221.716 365.385,221.716  
365.385,221.716 365.531,221.716 366.442,211.509 368.511,188.488 369.597,176.305"/>
```

```
</svg>
```

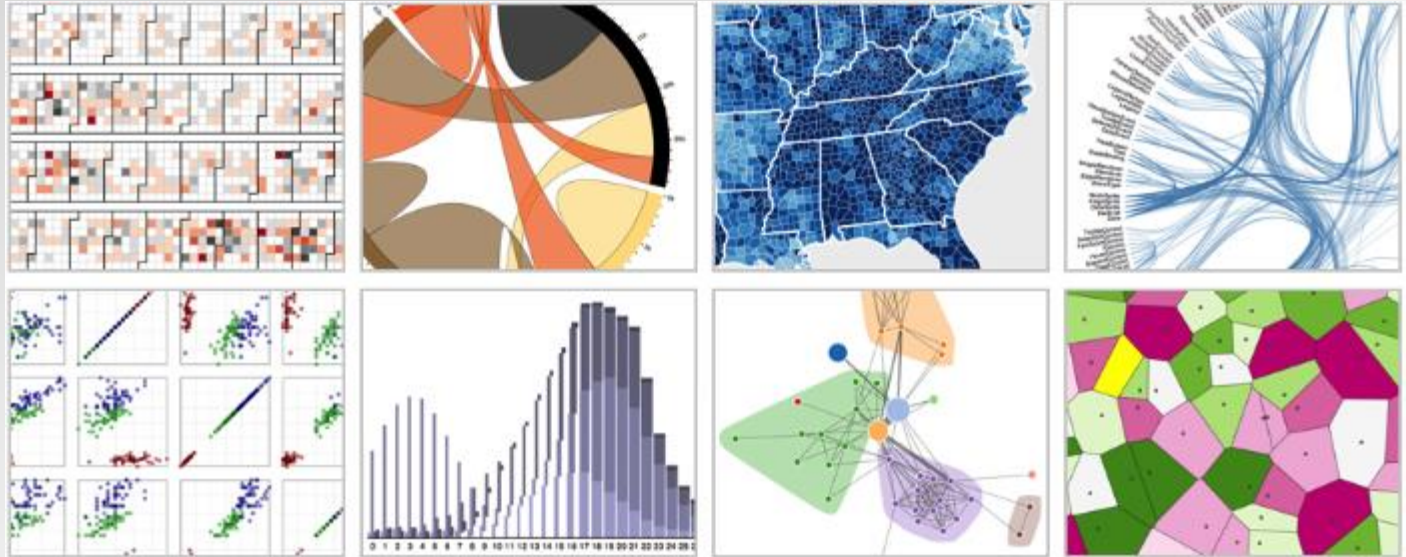


Source: <http://jsfiddle.net/danielfilho/GdCcA/>

# Plotting Charts with Google Charts

- Developed & Hosted by Google
- Closed-source, but free to use
- Uses SVG
- Very customizable
- Supports animations
- Example:  
<https://developers.google.com/chart/interactive/docs/gallery/columnchart#Example>
- Alternatives: flot, jqplot, HighCharts, etc. etc.

# d3.js



- [Source: http://d3js.org/](http://d3js.org/)
- **D3: Data-Driven Documents**  
Michael Bostock, Vadim Ogievetsky, Jeffrey Heer  
*IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011

# d3.js - Description

**D3** allows you to bind arbitrary data to a Document Object Model (DOM), and then apply data-driven transformations to the document. For example, you can use D3 to generate an HTML table from an array of numbers. Or, use the same data to create an interactive SVG bar chart with smooth transitions and interaction.

D3 is not a monolithic framework that seeks to provide every conceivable feature. Instead, D3 solves the crux of the problem: efficient manipulation of documents based on data. This avoids proprietary representation and affords extraordinary flexibility, exposing the full capabilities of web standards such as CSS3, HTML5 and SVG. With minimal overhead, D3 is extremely fast, supporting large datasets and dynamic behaviors for interaction and animation. D3's functional style allows code reuse through a diverse collection of [components](#) and [plugins](#).

Source: <http://d3js.org/>

# Using d3.js

- Though well-designed, has a steep learning curve
- Follows jQuery in style – provides operations on selections (e.g. everything can operate on multiple elements at once)
- Supports transitions
  - E.g., change this circle's size + position from 100 at (0, 0) to 150 at (45, 70) and make it move from the old to the new position in 1,000 ms.
- Requires OO JavaScript skills
- Fortunately: many components out there, many examples, very active user community.
  - E.g. WordCloud by Jason Davies [[Click for demo](#)]

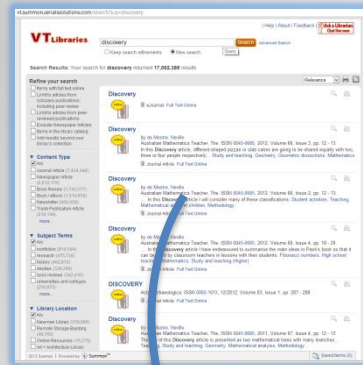
# Server Side Technologies Used

- As data changes, need to update client display
- Implemented in CoffeeScript using node.js [[Link](#)]
- Node.js: JavaScript platform for running server-side code
  - Based on Google's V8 JavaScript VM (same as used in Chrome browser)
- CoffeeScript: is a language that compiles to JavaScript, but uses a cleaner and richer syntax [[Link](#)]
- Socket.IO [[Link](#)]: is a library that supports client/server communication
  - Includes support for websockets
  - Kind of a successor to AJAX, makes "pushes" from server to client easier
- Code [is here](#)

# Introducing LibFX

- Right now, a bunch of scripts JavaScript, Python, CoffeeScript
  - Source code of course available
  - Contributions/collaborations are invited
- Idea: Build a cloud platform that allows other libraries to adopt and extend it
  - Particularly for non-programmers/non-sys admins
  - Not reliant on cloud, but ready to run “in the cloud” – on a IaaS or PaaS platform
- LibX Edition Builder experience: adoption really took off once we created a management interface that allowed independent setup & exploration

Users



Commercial or open source discovery system, e.g. Summon, EBSCO, Blacklight



Public Display & Interaction



Web-based visualizations, Incl. social media

libFX

Real-time Data Collection

Web-Based Control & Configuration Interface

Data Visualization

Data Processing & Storage

Offline Analysis

Augmentation

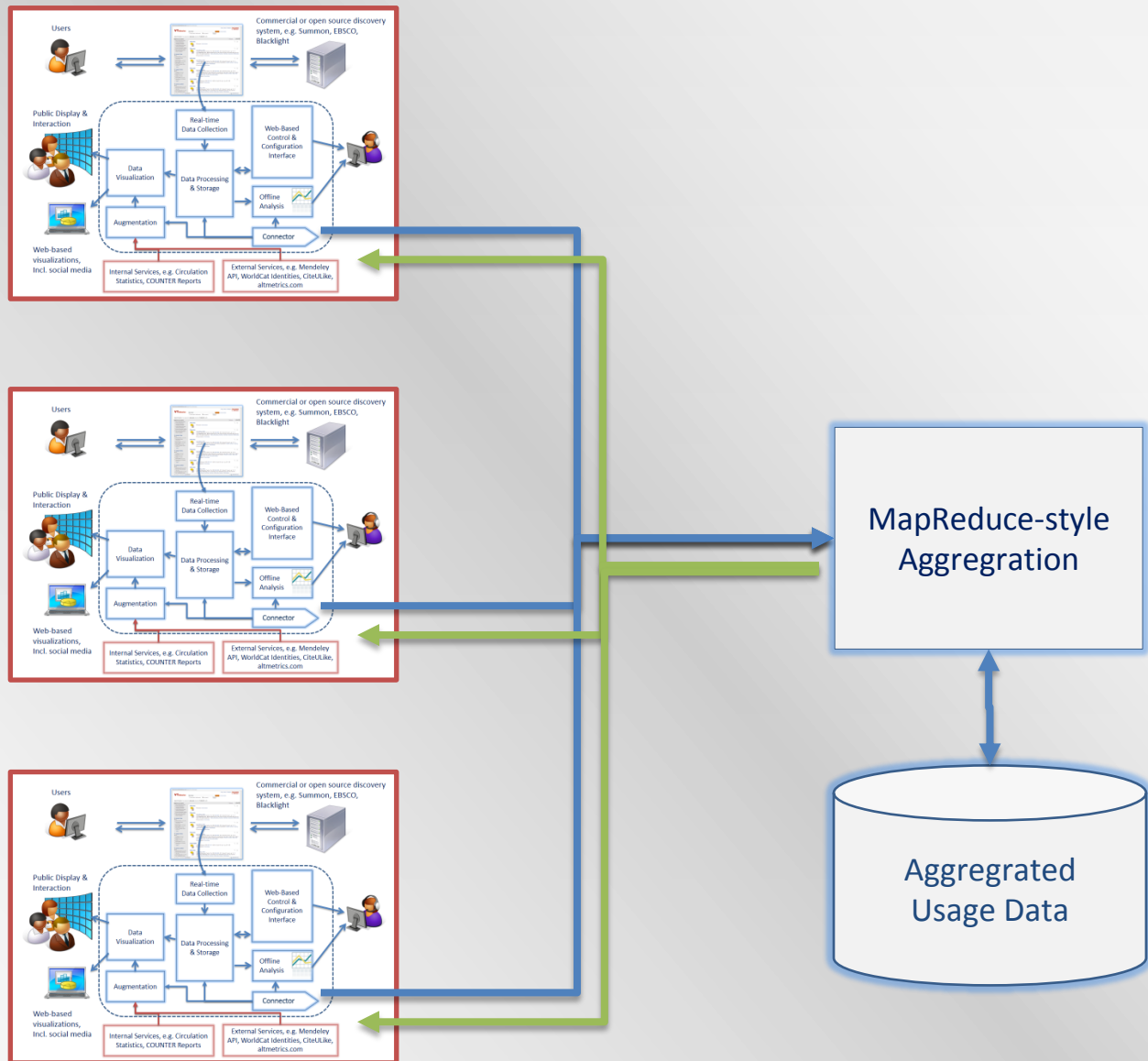
Connector

Internal Services, e.g. Circulation Statistics, COUNTER Reports

External Services, e.g. Mendeley API, WorldCat Identities, CiteULike, altmetrics.com







## Federation of libraries employing libFX

# Challenges & Opportunities

- What other information can be visualized?
  - Search terms? Search successes?
- How to create an attention-catching, yet unattended visualization
  - Digital Signage
- How to make the visualization interactive, allowing users to participate
  - E.g. QR codes, user feedback
- Last demo: [[Listen To Summon](#)]
- Website:
  - <http://libfx.lib.vt.edu>

## Credits

- Jason Davies for [d3 wordcloud](#)
- Bojan Đuričić for [justGage](#)
- [Stephen LaPorte](#) and [Mahmoud Hashemi](#) for [Listen To Wikipedia](#)
  - Based on Maximillian Laumeister [Listen To BitCoin](#)
- [The Art Of Web.com](#) for 3D Cube example

# Listen to Summon

