Authentication, Analytics and Gaining Insights for the Modern Library

By Scott R. Anderson
Associate Professor, Information Systems Librarian
Millersville University

scott.anderson@millersville.edu

LibTechConf 2019, Macalester College, Minneapolis, MN
Wednesday, March 20, 2019
What I said ...

• Modern authentication systems offer more security and simplified analytics, enable access to reporting, and classify users with customizable attributes for nearly endless reporting options. Learn about Millersville University’s rationale and process of migrating from IP / Proxy based access to utilizing OpenAthens authentication for access control and reporting. Library resources and services can now more readily meet the needs of local and remote university users and selected community patrons. Approaches to aggregating user behavior to make better-informed collection and resource decisions will also be discussed.
Got Questions?

Get my attention ....
and please ask.

(Just don’t throw anything at me.)
About Millersville University (MVS)

- An institution far far away …. 80 miles due west of Philadelphia
- Mid-Sized Comprehensive institution (6348 FTE)
  - 6779 (from 6778) undergrad, 1002 (from 970) graduate = 7781 headcount
  - 312 FT faculty, 180 PPT &/or TPT = 492 faculty
- 120+ programs for undergrads
- 24 programs for graduate, 3 Doc programs
- Top 10 declared majors: (1 to 10)
  - Biology, Business, English, Psychology, Math, History, Early Childhood Ed, CS, Meteorology, Early/Special Ed
- Several “dual institution programs” - EdD, MSW, DSW, MALC
About MVS Library

• Nearly 100% electronic
  • In 2015/16 - 5 physical journals; purchased 116 physicals
  • In 2016/17 - 4 physical journals; purchased 241 physicals (178 juvenile)
  • In 2017/18 - 4 physical journals; purchased 146 physicals (105 juvenile)
• 280k physicals and declining (general collections)
  • Maxed out just shy of 450k several years ago
  • 21k special collections; 22k juv/curr collections (reviewing this semester)
• 271k e-books and 170k e-journals/conference
  • JSTOR, EBSCO Completes, Muse, Sage, Elsevier, ACM, ACS, IEEE, IOP, Alex St Pr
• Previously - 20k e-books exposed via DDA (funds depleted)
• Previously - 34k e-videos exposed via DDA (funds depleted)
• Very robust Special Collections / Archives
  • Avg Sem: 250 students; 85 fac; 85 comm; 6k e downloads
About Scott R. Anderson (SRA)

• 2019 is 25th year in the library profession
• Instruction / Research assistance
  • Art & Design, Business, Computer Science, Economics, Mathematics
• Represent entire library systems effort
  • ILS, ILL, IR, SC&A, content platforms, Discovery, Linking, KB, ... and user access
    • Alma, Illiad, TIND IR, Aeon & ASpace, EDS, FTF ... AD to Shib to Proxy/Shib/OA
• Collections use and analysis
  • Local and consortium collections/resources
  • Analysis of use - subscribe vs. buy vs. borrow vs. cost per use
  • Value propositions - renew, not-renew, expand/shrink versioned resources, switch
  • Ways to meet curricular changes
Access: users, services, content

- **On Campus: Previously IP Access**
  - Easy, walk up and go .... Almost ... university login now required
  - Someone used a resource - not as helpful for analysis – no user attributes
  - Not the same experience as “off campus”- focus group results surprised us

- **Off Campus: Previously Proxy**
  - Consortium run authentication / proxy / Shibboleth-ish / **customized**
    - Carrying the costs of customization in staff time and maintenance
  - Y/N authentication – not as helpful for analysis – no user attributes
  - Reporting never a high priority

- **No ILS “back door” access mechanism for years**
  - Avoid “patron file” – use real time University IdP when possible
Access: **users, services, content**

- University is official identity provider
  - Actionable to the library via some IT service
  - Official source of basic student/faculty/staff “user” and “group” types of data
  - Making attributes actionable in 3 steps….
    
    (1) Student Information System (SIS), then to  
    (2) Active Directory (AD), then to  
    (3) Shibboleth (SSO) so we can use it….
    
    … Is automated … or it seems to be to us anyway ….

- Getting SPECIFIC attribute data flowing is in testing: major, minor, courses
Access: users, services, content

- IT provides the basic attributes very well
  - Provides in an actionable way the basic attributes
    - Username / email / group affiliation / current status
    - memberOf “faculty / staff / student” ... and “currently a student / employed”

- IT provides the more specific attributes .... not quite yet
  - Previously : You want what? ... go ask the Registrar
  - Specific attributes automated this Spring / Summer
    - Think “course” or “major” or “minor” or “dept” or “research assistant” or “credit hours”
    - Big distinction between “single value” and “multiple value” data points
    - Attributes that would assist with scoping access to specific resources / services

- Attributes allow us to manage logic to ...
  - scope access / get more informative reports
Access: users, services, content

- InterLibrary Loan
  - Illiad - user provides one time demographics, doesn’t update ... yet (next on roadmap)
  - Illiad - working with OA since Feb 26, 2019

- Course reserve
  - Ares – one time course affiliation per semester ... will become auto-populated

- “E-Zborrow” for unmediated monographic requesting (OCLC Relais -> ReShare)
  - Uses that icky patron file from the ILS
  - What’s my Library ID number?

- Repository
  - Submission permissions (if attribute “submit”, you get permission)

- Discovery
  - Real time ACCOUNTS being created in Discovery via OA. Previously via Shib
  - Accounts are UNIQUE into “perpetuity” so are perfect User ID for Millersville
  - Building out personalization around this UI – pull ”personalization” from other services
Access: users, services, content

• Very “e content centric”
  • Seamless access is a big issue – because it’s basically all our content
  • We buy physicals as a last resort

• Flat budgets
  • Analysis of content “costs” and “value” increasingly important
    • Value proposition “varies” by user group
  • To add/alter content mix, we need to curtail, trim, cut, or slash
  • Clearly demonstrate ability to meet niche need to pry out non-library $$$
    • We think “scoping access” to resources will help –
    • Example: “content for course by semester” ... charged back to the department ( ? )
Midway Brief Recap

• 100% electronic as a practical matter

• Moving away from IP / Proxy to secure and “known” SSO process
  - Avoid: Lack of security, no reporting, no scoping, IT managing 125+ connections
  - Gain: Security w/SAML, actionable reporting, scoped access, ONE connection

• Library gains a singular mechanism to manage users to content
  - Leverages what the University IdP provides
  - Provides Library w/options to access when lacking university provided attributes

• IT manages a single connection to OpenAthens
  - Ie, the Library doesn’t need to bother them about connections!

• Users gain a singular known process to login regardless of location
What is OpenAthens? How does it work?

• A “gateway authentication” service
  • Most libraries have a “general scholarly research database”
  • Metaphorically similar, but for authentication with content providers
  • OA and content providers “talk” via a set of commonly understood attributes

• Connect your institution to OpenAthens
  • IT makes this connection one time (via some SAML service)
  • Determine the universe of user attributes that you want to release to OA
  • Connect your OA profile to your content providers and services
    • Release attributes to each content provider or service as appropriate
  • Leverage existing channels and “WAYF” if user on open web
Advantage of OA for the Library?

• Catalogue of connections
  • The library can handle connections – add, drop, etc
  • Single point to manage/configure connectivity

• Additive functionality
  • Access control logic independent of what IT has implemented
    • IT can’t tell us that, but the Registrar can
    • Non-university use cases the library must address

• Actionable Reporting
  • Even if local IT wanted to make all the connections
  • They probably can’t do reporting
    • We could do just InCommon if we had the time;
    • But we can’t readily tell you about the activity.
Reporting & Analysis with OA

• Easy to generate reports
  • Basic to super-fancy
  • Richness based on what attributes you have mapped
  • More attributes or attribute values, more reporting options

• Drop into Excel / Reporting tool of choice
  • Commence numerical wizardry
  • Combine OA reports with other data
The reporting “dashboard” quick view

<table>
<thead>
<tr>
<th>Top resources</th>
<th>Last 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCO Information Services</td>
<td>16,154</td>
</tr>
<tr>
<td>ITHAKA</td>
<td>1,912</td>
</tr>
<tr>
<td>Elsevier</td>
<td>1,019</td>
</tr>
<tr>
<td>Atlas / ILLiad</td>
<td>756</td>
</tr>
<tr>
<td>Wiley Online Library</td>
<td>518</td>
</tr>
</tbody>
</table>

**Total authentications**

11,373

**Total accounts**

6,877

- Active accounts: 6,874
- Expired accounts: 0
What do the reports say? ... about users

• 6,780 Millersville library resource/service since 8/1/2018
  • 1.07 accounts for every FTE; or 82% of all students & faculty by headcount
  • That’s far more than we knew before

• Dual institution programs:
  • 11 Kutztown, 36 Shippensburg, 9 West Chester
  • Again, we had no idea if this was even really working previously

• Local HS students (“personal” OA accounts)
  • 44 accounts created; 41 accounts are active
What do the reports say? .... about resources

• EBSCO is our most popular “platform”
  • This is no surprise given our content / service mix
  • What is surprising is the degree to which that is the case
    • Biggest by a factor of about 9 – but we don’t have a full year in yet

• “Redirections”
  • Keep in mind, these are “redirection requests into the PLATFORM”
    • Still need to make use of “counter like” reports for volume of activity ON a platform
      • Your FT requests, sessions, link-ins, link-outs, etc.
      • COUNTER reports still matter
**Scope:** This organisation
**Resource:** All resources
**Breakdown by:** Resource
**Granularity:** Monthly
**Start:** 1 August 2018
**End:** 1 March 2019

**Chart Description:**
- The chart shows a trend from 1 August 2018 to 1 March 2019.
- The y-axis represents a scale from 0 to 20,000.
- The x-axis represents dates from 01 Aug 2018 to 01 Mar 2019.
- The chart includes data points for several resources.

**Resource Breakdown:**
- **EBSCO Information Services:** 65,132, average 8,141.5, min/max 0/17,663
- **ITHAKA:** 6,496, average 812, min/max 0/2,106
- **Elsevier:** 3,508, average 438.5, min/max 0/878
- **CQ Researcher (IP):** 2,135, average 266.9, min/max 0/899
- **SAGE Journals:** 1,578, average 197.3, min/max 0/463

**Legend:**
- Red: EBSCO Information Services
- Blue: ITHAKA
- Green: Elsevier
- Orange: CQ Researcher (IP)
- Purple: SAGE Journals

**Change:**
- EBSCO Information Services: 2710
- ITHAKA: 113
- Elsevier: 147
- CQ Researcher (IP): 213
- SAGE Journals: 80
<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>TOTAL</th>
<th>AVERAGE</th>
<th>MIN / MAX</th>
<th>CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCO Information Services</td>
<td>58,644</td>
<td>11,728.8</td>
<td>2710 / 17663</td>
<td>↑ 6216</td>
</tr>
<tr>
<td>ITHAKA</td>
<td>5,823</td>
<td>1,164.6</td>
<td>113 / 2106</td>
<td>↑ 942</td>
</tr>
<tr>
<td>Elsevier</td>
<td>3,027</td>
<td>605.4</td>
<td>147 / 878</td>
<td>↑ 404</td>
</tr>
<tr>
<td>CQ Researcher (IP)</td>
<td>2,027</td>
<td>405.4</td>
<td>213 / 899</td>
<td>↑ 23</td>
</tr>
<tr>
<td>SAGE Journals</td>
<td>1,455</td>
<td>291</td>
<td>80 / 463</td>
<td>↑ 98</td>
</tr>
<tr>
<td>RESOURCE</td>
<td>TOTAL</td>
<td>AVERAGE</td>
<td>MIN / MAX</td>
<td>CHANGE</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------</td>
<td>---------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>EBSCO Information Services</td>
<td>16,154</td>
<td>521.1</td>
<td>0 / 837</td>
<td>↓749</td>
</tr>
<tr>
<td>ITHAKA</td>
<td>1,912</td>
<td>61.7</td>
<td>0 / 112</td>
<td>↓61</td>
</tr>
<tr>
<td>Elsevier</td>
<td>1,019</td>
<td>32.9</td>
<td>0 / 74</td>
<td>↓59</td>
</tr>
<tr>
<td>Atlas / ILLiad</td>
<td>756</td>
<td>24.4</td>
<td>0 / 95</td>
<td>-</td>
</tr>
<tr>
<td>Wiley Online Library</td>
<td>518</td>
<td>16.7</td>
<td>0 / 41</td>
<td>↓8</td>
</tr>
</tbody>
</table>
What do the reports say? ... about a service?

• Combining ILL and OA reporting to gaining some preliminary insights

• OA Redirect reporting
  • “in aggregate”
  • “by user type”

• Compare / contrast that to volume of activity of a service
  • Example: Illiad requests
  • “in aggregate”
  • Could break down “by user type”;
    • but that would be manual lookup in Illiad at this time
<table>
<thead>
<tr>
<th>EDUPERSONSCOPEDAFFILIATION</th>
<th>↓ TOTAL</th>
<th>AVERAGE</th>
<th>MIN / MAX</th>
<th>CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:student@millersville.edu">student@millersville.edu</a></td>
<td>669</td>
<td>21.6</td>
<td>0 / 87</td>
<td>-</td>
</tr>
<tr>
<td><a href="mailto:faculty@millersville.edu">faculty@millersville.edu</a></td>
<td>77</td>
<td>2.5</td>
<td>0 / 11</td>
<td>-</td>
</tr>
<tr>
<td><a href="mailto:staff@millersville.edu">staff@millersville.edu</a></td>
<td>10</td>
<td>0.3</td>
<td>0 / 3</td>
<td>-</td>
</tr>
<tr>
<td>Date</td>
<td>Totals</td>
<td>faculty</td>
<td>staff</td>
<td>student</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Total</td>
<td>731</td>
<td>75</td>
<td>10</td>
<td>646</td>
</tr>
<tr>
<td>2/27/19</td>
<td>50</td>
<td>8</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td>2/28/19</td>
<td>63</td>
<td>7</td>
<td>3</td>
<td>53</td>
</tr>
<tr>
<td>3/1/19</td>
<td>39</td>
<td>7</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>3/2/19</td>
<td>33</td>
<td>2</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>3/3/19</td>
<td>33</td>
<td>4</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>3/4/19</td>
<td>75</td>
<td>3</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td>3/5/19</td>
<td>87</td>
<td>11</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>3/6/19</td>
<td>69</td>
<td>4</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>3/7/19</td>
<td>95</td>
<td>5</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>3/8/19</td>
<td>52</td>
<td>9</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>3/9/19</td>
<td>23</td>
<td>3</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>3/10/19</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>3/11/19</td>
<td>36</td>
<td>4</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>3/12/19</td>
<td>27</td>
<td>6</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>3/13/19</td>
<td>32</td>
<td>2</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

Requests per Redirect
And some student assessment? ...

- Option to allow capture of “redirects into a platform” by user
  - OPTIONAL
  - Use whatever data element you want
    - Username, e-mail, **student ID number**, something that is being passed
      - Can’t tell what they “did” on a platform
      - Only that they were redirected into a platform

- Some examples with scrubbed data suitable for presentation
### OpenAthens statistics by username/UID broken down by resources

**Administrator:** XXXX

**Report period:** 01 January 2018 - 26 February 2018

**Date run:** 26 February 2018 02:22:12 GMT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>10302</td>
<td>163</td>
<td>42</td>
<td>27</td>
<td>1</td>
<td>2</td>
<td>293</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Total Access by "Username", Jan & Feb 2018

Possibilities:
- Assessment
- Surveys
- Focus Groups
- "User Persona"
Key Opportunities for MVS

• ‘Known SSO” login process regardless of location
  • On or off campus; across devices; users experience the same process
  • Portability between resources / services, “single sign on”

• Lower IT hurdle
  • IT admits they can’t realistically manage 125+ connections
  • They can manage one … OpenAthens
  • Focus on making user attributes more readily available
  • Might even use OA as the University SAML service vs. Shibboleth

• Reporting
  • What groups are using what resources and to what extent
    • Is use reasonable? Is there sufficient value ?
Key Opportunities for MVS

• Collections Analysis / Scoping
  • Program proclaims “we must have”; but your students are not using ....
    • The Library can help with that ...
  • ”Right size” access to content with users of that content
    • By program, course, length of engagement

• Personalization
  • Platform accounts without “making another account”
  • Working on pulling that all together in our “library UI”
    • This might turn out to be “Stacks” or something akin to that

• Security
  • Best practice
Thank you

Got Questions?

scott.anderson@millersville.edu