Shared Repository Infrastructure
CNI Fall 2017 Membership Meeting

- Jimmy Ghraphery, Associate University Librarian for Scholarly Communications and Publishing, VCU Libraries
- jghapher@vcu.edu | @jimmyghaphery
Measuring our full Elsevier investment

• VCU bepress subscription 2.2% of total Elsevier serial spend
• 26% of VCU collections budget Elsevier and Wiley serial
• VA Elsevier Consortium Total Spend 2017-2021 $45.47 million
• Does any university know how much? (monographs, textbooks, labor, etc)

2.5% corollary: Vendor subscriptions are an investment in commercial infrastructure
The bepress 115

- 115 “Research Universities” using Digital Commons (DC)
- Publicly available data collected 9/28/2017 (apples::apples)
- 115 of 452 (25%) of all DC institutions
- Est. 60% downloads 75% of items of all DC sites
The bepress 115

- Carnegie Classification
- Journal publishing
- ETDs (map data, qualitative notes and quotes)
- Dataset on OSF (DOI: 10.17605/OSF.IO/2KPG6)
bepress 115 Carnegie Class

- Doctoral Highest: 28.7%
- Doctoral Higher: 37.4%
- Doctoral Moderate: 23.5%
- Master's Colleges & Universities: 7.8%
- Canadian: 7.8%
bepress 115 Carnegie Class

- Doctoral Highest: 37.4%
- Doctoral Higher: 28.7%
- Doctoral Moderate: 7.8%
- Master's Colleges & Universities: 23.5%
- Canadian: 2.7%
bepress 115 and Carnegie Docs

- Doctoral Highest: 115
- Doctoral Higher: 107
- Doctoral Moderate: 112

- bepress 115
- Carnegie
bepress 115 and Carnegie Docs

31% of Carnegie Docs
bepress 115 Age (downloads/downloads past year)

bepress 115 avg: 2.81

4 years?

VCU 1.8 = 3 years
bepress 115 Size and Downloads

- 1.5 million items
- 247 million total downloads
- 62 million downloads in past year
- Top 10: .5 million items 23 million annual downloads
bepress 115 Size and Downloads

- 1.5 million items
- 247 million total downloads
- 62 million downloads in past year
- Top 10: .5 million items 23 million annual downloads

*Bepress currently has content containing over 2 million articles and 100 million annual downloads. (Press Release Aug 2, 2017)*
bepress 115 Size and Downloads

• 1.5 million items
• 247 million total downloads
• 62 million downloads in past year
• Top 10: .5 million items 23 million annual downloads

Bepress currently has content containing over 2 million articles and 100 million annual downloads. (Press Release Aug 2, 2017)

A lot or a little?
Infrastructure: Stats (apples::oranges)

- BePress items: every published record = 1 item, includes metadata-only record and record with multiple supplemental files.
- Dspace/Samvera: count from browse all titles (similar to bepress definition)
- Stanford/Harvard/MIT: published stats
- Vendors: public stats, many self reported
University Infrastructure: Size

<table>
<thead>
<tr>
<th>Repository</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carolina Digital Repository</td>
<td>266,974</td>
</tr>
<tr>
<td>University of California eScholarship</td>
<td>154,294</td>
</tr>
<tr>
<td>University of Michigan DeepBlue Docs</td>
<td>115,086</td>
</tr>
<tr>
<td>University of Illinois IDEALS</td>
<td>93,973</td>
</tr>
<tr>
<td>MIT Dspace@MIT</td>
<td>87,148</td>
</tr>
<tr>
<td>Texas A&amp;M OAKTrust</td>
<td>81,924</td>
</tr>
<tr>
<td>Texas ScholarWorks (UT)</td>
<td>52,814</td>
</tr>
<tr>
<td>Cornell eCommons</td>
<td>47,032</td>
</tr>
<tr>
<td>Harvard Dash</td>
<td>44,046</td>
</tr>
<tr>
<td>Columbia Academic Commons</td>
<td>23,660</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>966,951</strong></td>
</tr>
</tbody>
</table>
Infrastructure: Annual Visits

Visits (millions)

- ResearchGate: 600.0
- ScienceDirect: 400.0
- Wiley Online: 200.0
- SpringerLink: 100.0
- Sci-Hub: 50.0
- bepress: 25.0
- bepress 11.5: 10.0

bepress=downloads
Infrastructure: Visits per item

- ResearchGate
- ScienceDirect
- Wiley Online
- SpringerLink
- Sci-Hub
- bepress
- bepress 115

Visits/Item
Bepress 115: ETDs

Items

- ETDs: 15.7%
- Not ETDs: 84.3%

Downloads past year

- ETDs: 31.8%
- Not ETDs: 68.2%

Harvard: 9% items | 20% downloads
ETDs 2:1 rule – Why?

Harvard: 9% items | 20% downloads
ETDs and infrastructure so what

- How Open is it?
- Multiple/Model license considerations
- Education opportunity
- Submission workflow
- National ETD preservation platform “value content transfer”
- Is 15-30% of repository development centered on ETDs?
- OSF for ETDs, Vireo, COAR Next Generation Repositories, ETD community
The “1 thing” Digital Commons gets right

- Shared publishing/repository
- Readership maps
- Author email alerts
- Faculty profiles
- SaaS
- Others..............................
Commercial Goal: Customer Success

- Help and support
- Community of users (platform to policy)
- Backend UI matters
- Broad conception of repository
- Governance
Federated Scale

- Standardization Any color Model T
- Enhancements
- No size or journal limits
- Aggregation
- Benchmarking analytics
- Leverage cloud backups
- Different expectations from in house/subscription services
Credits and References

• Bepress 115 data set. OSF. DOI: 10.17605/OSF.IO/2KPG6
• COAR Next Generation Repositories Working Group, Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group (2017.11.28)
• Harrington, Robert . The Scholarly Kitchen. ResearchGate: Publishers Take Formal Steps to Force Copyright Compliance (2017.10.06)
• Harvard DASH Stats (2017.10.06)
• Images Nasa Globe and USDA Broccoli
• MIT Open Access Article Statistics (2017.10.02)
• Schonfeld and Ulmschneider. ASERL Webinar. “Red Light, Green Light: Brainstorming how ASERL members might-could test a new licensing model” (2017.10.13)
• Stanford Digital Repository Overview
• Thesis Commons (OSF)
• Vireo ETD
Thanks and questions

Jimmy Ghaphery

jghapher@vcu.edu