




# Coalescing Usage Data

---

Research, Data Driven Decisions &  
User Privacy



# What are we talking about?



Risk to User Privacy



Circulation Data Retention Trends



Project to Lower Risk & Enable Library Research



Aggregated Data & Use



Challenges & Opportunities

# Academic Freedom at Risk

# Targeting a Climate Scientist

- April 2010 VA Attorney General filed a subpoena seeking documentation of Michael Mann's time at UVA and launched an investigation
- Many organizations and individuals came out in support of academic freedom in this case
- August 2010 a VA Circuit Court dismissed the investigation finding no factual basis
- 2011 The American Tradition Institute (a free market think tank) followed this dismissal with a Freedom of Information Act (FOIA) request
- In 2014 The Supreme Court of VA unanimously affirmed the earlier Circuit Court decision and ruled in the University's favor

# The Phoenix Project

- June 2015, a Federal agency notified UVA's CIO that "two nation state actors may have access to the university's systems"
- 176 staff, one government agency, and two commercial companies commenced to determine the extent of the compromise and devise a plan for remediation.
- August 2015, UVA "goes dark" for remediation
- UVA learned that two faculty were under surveillance by a nation state

# Library Vendor Surveillance

- 2020 Libraries published research on the growing problem of online surveillance, and censorship
- April 2021 RELX a subsidiary to LexisNexis signed a multi-million dollar contract with US Immigration and Customs Enforcement Agency (ICE)
- The ICE contract revealed that LexisNexis' databases would "offer an oceanic computerized view of a person's existence" and would provide the agency with "the data it needs to locate people with little or no oversight"
- Both RELX and Thomson Reuters have built global surveillance systems that include online tracking, they have a massive growing aggregation of user data, and sell of services based on that tracking

# Patriot Act

- October 2001, Patriot Act was established.
- February 2010, the Patriot Act was reauthorized without reform regardless of documented abuse
- August 2019, the National Director of Intelligence called on Congress to reauthorization based on “no history of abuse for 18 years”
- March 2020, an effort to extend did not pass the House of Representatives and therefore the law expired.
- So why is this important to this discussion?

# Circulation Data at UVA

## What data are at risk?

- University ID
- Computing ID
- Over a decade of data

## Why?

- Locate lost items
- Prove library usage to schools questioning “library tax”
- Inform collection building through usage trends
- Validate surveys (or try)
- Staffing

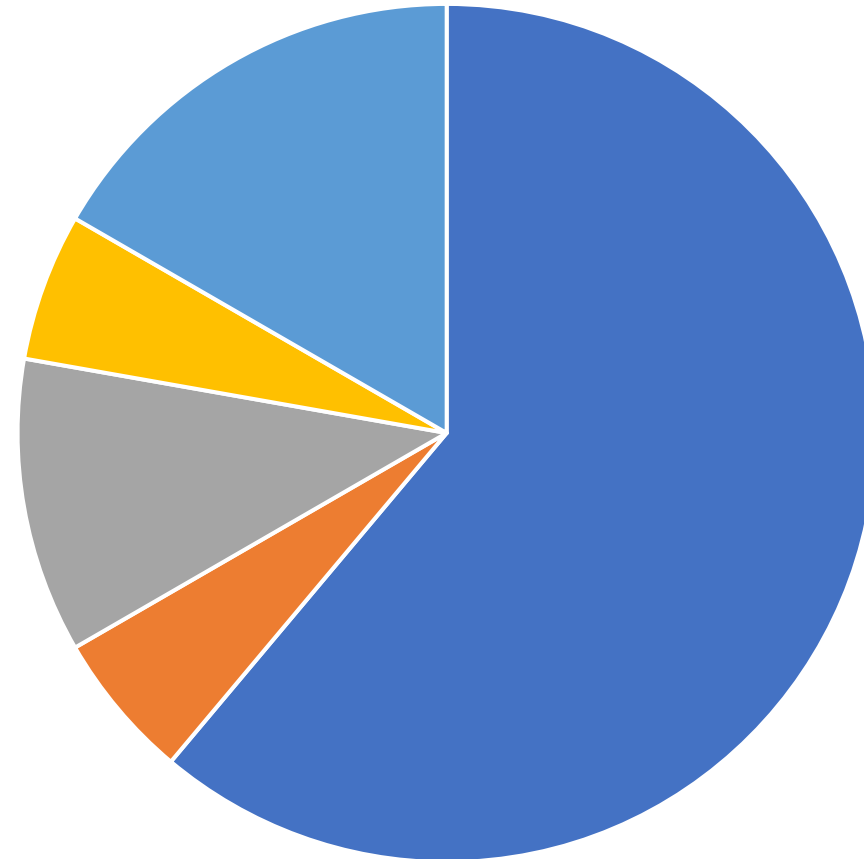
## Our Primary Risk?

- Court Orders
- Security breaches



# Circulation Data Trends

Retention



■ 0-1 month ■ 6 months ■ 1 year ■ 3 years ■ Unlimited

# UVA Project Goals

## Reduce

Reduce Risk to User  
Privacy

## Augment

Augment Data with  
Disparate Library &  
Enterprise Systems

## Enable

Provide Search &  
Access Enabling  
Research & Data  
Driven Decision  
Making

# Collaboration & Impact

## Library Senior Leadership

## Institutional Partners

- UVA Institution Research & Assessment (IRA)
- UVA Institution Review Board (IRB)
- UVA Information, Technology Services (ITS) & CISO

## Resulting Approach

- Revised Retention & Archiving Schedule
- De-identification
- Security & Access

# Aggregate Data

- Time/Date
- School
- Department
- Unique Key
- Job Family Group
- Job Family Name
- Roles
- Borrower Profiles
- Faculty Type
- Degree Seeking
- Degree Level
- User Library
- Station Library
- Item Library
- Item Location
- Reserve Desk
- Item Data
- Primary Subject & Subject Fields
- Item Class Scheme ( e.g. LC, SUDOC)
- Item Type (e.g. Book, CD, Equipment)
- Format ( e.g. Monograph, Serial, Map)
- Publication Date
- Language

# Search, Mine, Extract

- Make available all possible values
- Search by any combination of selected attributes and values
- Browse results to validate search
- Sort by any attribute
- Export data for further processing

# Technical Decisions

- Automate to Meet Retention Schedule
- Cloud Infrastructure
- SOLR Index for Speed & Flexibility
- Export CSV for Post Processing
- Archive Source Files

# Library Assessment Usage

- Access to Living Data – aggregated across systems
- Conduct Logitudinal Analysis
- Validate Surveys & Bust Myths

“Commerce  
doesn't use  
the Library”

- myth
- circulation consistent over a decade, mostly books and DVDs
- most heavily used library- Clemons
- most common borrower profile – undergrad



What  
classifications  
at a high level  
are faculty  
using?

- By department, which faculty use the Library most often? English & History
- Library most used by faculty? Main Library, but also the beloved LEO service
- The “Ps” as expected had the highest circulation with English Literature as the highest sub classification



## Informing Library Collections

- Who really uses what?
- Usage patterns vs anecdotal feedback
- Immediate and long-term analysis of collection use
- Collections Management and Maintenance Decisions
- Framework for good stewardship
- Can we correlate user success to collection use?

# Does interdisciplinary or shifting research impact collection usage? E.g. Archeology

- Expected faculty use from the Art Department where Archeology is taught or Anthropology
- Learned that History usage was far greater
- Student usage was double that of faculty
- Most materials are circulated between 1 and 2 PM





## Who uses the Arabic Collection?

- Overhauled our Arabic Collection Plans
- Learned that the School of Continuing and Professional Studies also heavily uses the collection
- Learned that use overall was evenly split across faculty, graduates and undergraduates

# Research & Learning Services

- Understanding use patterns among departments, subject areas, and user groups we work with
- Verifying anecdotal evidence / trends
  - E.g. undergraduates saying they want "fun fiction"
- Identifying collections areas for outreach and development
- Identifying areas for instruction
  - E.g. building tutorials for finding books in a specific location or call number range
- Can help DEI efforts – who is using our resources? Where can we build collections?

# Potential Use on the Front Lines



LOCATION USAGE  
TRENDS



ITEM LOCATION VS  
CIRCULATION LOCATION



CIRCULATION TRENDS BY  
DATE/TIME

# Challenges Encountered

- Diversity, Equity, and Inclusion Data
- System & Data Migrations Impacting Data Quality
- “Weeding the Collection” Gaps
- Increase of Digital Content & Authenticated Access Resistance

Opportunities?

Compare &  
Contrast?

Aggregate  
Across  
Institutions?



## Acknowledgements

---

Beth Blanton

---

Lou Foster

---

Bob Gartland

---

Dave Goldstein

---

Bob Haschart

---

Tim Morton

---

Cecelia Parks

---

Annette Stalnaker

---

Penny White

---

# Credits

- “Timeline: Legal Harassment of Climate Scientist Michael Mann.” *Union of Concerned Scientists*, <https://www.ucsusa.org/resources/legal-harassment-michael-mann>.
- Tanczer, L. M., Deibert, R. J., Bigo, D., Franklin, M.I., Melgaco, L., Lion, D., Kazansky, B., & Milan, S. “Online Surveillance, Censorship, and Encryption in Academia.” *International Studies Perspectives*, 2019, <https://doi.org/10.1093/isp/ekz016>
- Nelson, R., Wright, R. *The Phoenix Project: Remediation of a Cybersecurity Crisis at the University of Virginia*. Darden Business Publishing, September 27, 2017
- “Addressing the Alarming Systems of Surveillance Built by Library Vendors.” *SPARC*, 12 Apr. 2021, <https://sparcopen.org/news/2021/addressing-the-alarming-systems-of-surveillance-built-by-library-vendors/>.
- Lynch, Clifford. “The Rise of Reading Analytics and the Emerging Calculus of Reader Privacy in the Digital World.” *First Monday*, vol. 22, no. 4, 2017, <https://doi.org/10.5210/fm.v22i4.7414>.

---

# Thank you!

---

Robin Ruggaber

[robin.ruggaber@virginia.edu](mailto:robin.ruggaber@virginia.edu)