PRIVACY AND THE RESEARCH LIBRARY: PERSPECTIVES FROM THE US

Helen Cullyer
Program Officer, Scholarly Communications
The Andrew W. Mellon Foundation

PRIVACY, TECHNOLOGY, AND POLITICS

New technologies provide new ways for criminals, governments, and corporations to collect, analyze, and exploit factual and behavioral information about individuals.

The right to privacy is often challenged during wars and in the face of threats to the socio-political status quo.

A BRIEF HISTORY OF PRIVACY IN THE US

- Brandeis and Warren (1890): “instantaneous photographs” and “the newspaper enterprise”.
- Censorship and requests for patron records during the First World War.
- Censorship and surveillance during the “Red Scare” and Prohibition eras. Federal wiretapping.
- 1939: American Library Association (ALA) code of ethics espouses a commitment to protecting the privacy of library patron records.
- Further governmental challenges to librarians’ code of ethics: McCarthyism during 1950s and Patriot Act of 2001.
- 1990s-present: Data breaches at major universities, including University of Maryland and Stanford University.

DISCOVERY AND THE ONLINE CATALOG: SEARCH, BROWSE, AND SHARE

Primo

BIBLIOCOMMONS

Summon™
Web Scale Discovery

vufind

blacklight

WorldCat®

CURATION, SHARING, REVIEWING, TAGGING, RANKING, AND RECOMMENDER SERVICES

- Are privacy settings transparent?
- What information that is personally identifiable is the library, and / or vendor hosting the system, tracking and collecting?
- How long is library / vendor retaining that information?
- To what third parties are data disclosed?
- Are usernames and passwords stored and transmitted in encrypted formats?
- How easy is it to interpret the library's the vendor's privacy policies?

ERIC HELLMAN:

ANALYSIS OF PRIVACY LEAKAGE ON LIBRARY CATALOG WEBPAGE

> “In building the *Communist Manifesto* catalog page, my browser contacts 11 different hosts from 8 different companies.”

http://go-to-hellman.blogspot.com/2014/09/analysis-of-privacy-leakage-on-library.html

HTTP vs. HTTPS

HTTPS:

- Authenticates the website and web server with which one is communicating
- Provides bidirectional encryption of all communications

When an https page links to an http page, the request does not include the “referer” header: i.e. the URL of the https page from which the request originated is not revealed. This provides some, albeit limited, protection against tracking browsing history.

ACCESSING DIGITAL CONTENT

…using the slightly anachronistic example of the *Communist Party Manifesto*

SCENARIO ONE:

A research library has digitized the first English-language edition of the Communist Party Manifesto and provides digital access on open access basis (to library patrons and others on the open web).

- What are the privacy policies of the library and of any third-party hosting the content?
- Are third-party social sharing services available? Is so what data do those services collect and track?
- Is content available via http or https?
- To what extent does library and / or third-party track actual reading behavior?
- If users outside the community of library patrons are asked to create username and password to access the content, how are those data transmitted and stored (encrypted or plain text)?
- Are users able to annotate content? If so, what authentication mechanisms are employed for users outside the community of library patrons?

SCENARIO TWO:

A library provides access to an open access version of the Communist Party Manifesto, hosted by a publisher, or aggregator, or other content provider.

Privacy issues are the same as in scenario 1, except that the relevant policies and mechanisms are those of the publisher, aggregator, and other content provide.

SCENARIO THREE:

A library provides access, to authenticated users only, to a modern critical edition of the Communist Party Manifesto, hosted by an aggregator or publisher, behind a paywall.

Specific privacy and security concerns depend on the method of access and authentication:

- Direct IP-based access on internal campus / library network
- Remote Access and Authentication
    1. Proxy Server
    2. Virtual Private Network
    3. Trusted federated authentication networks, such as Shibboleth and OpenAthens.
    4. Many others…

SCENARIO FOUR:

The library hosts a digital humanities project that involves collaborative creation of a multi-lingual, aligned, annotated critical edition of the Communist Party Manifesto.

- What are the privacy requirements? Are all comments and contributions open? Are all contributors identified by real names?
- What are access and authentication mechanisms?
- http vs. https
- What, if any, third-party services are integrated?

PRELIMINARY CONCLUSIONS

- Patron data now means not just data *about* patrons but also data contributed *by* patrons (to the catalog or to digital resources themselves)
- In addition to patron data, libraries need to think more broadly about user data (i.e. users who may not hold library cards and IDs, but interact with open access resources provided by a library).
- Libraries are no longer solely in control of patron and user data. Publishers, aggregators, and software vendors play an equally important role.
- Patron / user data are useful data.
- The very services that make the modern library catalog and digital content interactive, engaging and more useful are those that may endanger user privacy.

LIBRARIES AND PRIVACY IN THE US

Two privacy literacy initiatives, funded by Knight Foundation: http://www.knightfoundation.org/blogs/knightblog/2015/1/30/22-projects-win-knight-news-challenge-libraries/

Ongoing advocacy and research by the American Library Association: http://www.ala.org/advocacy/privacyconfidentiality

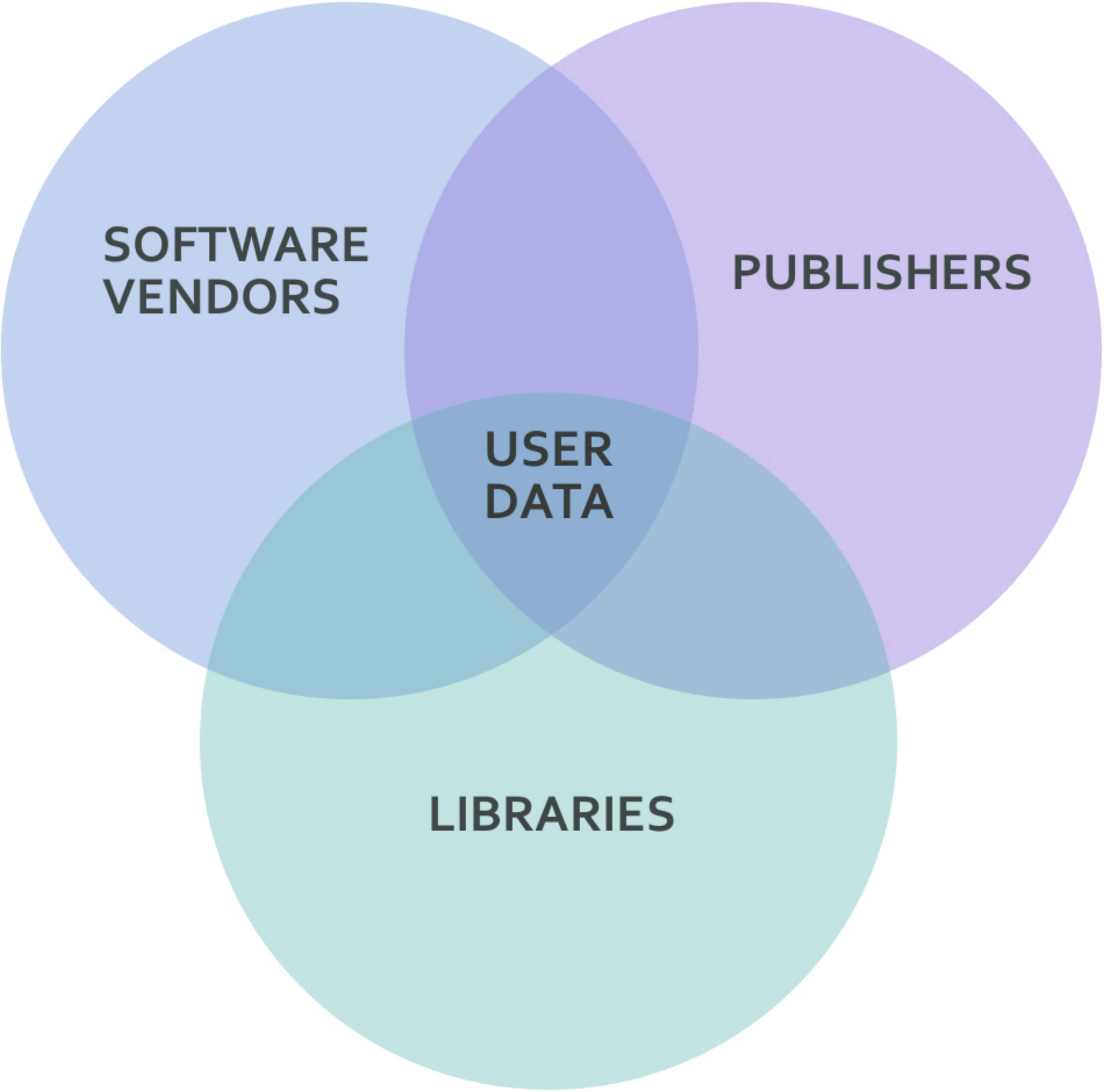
University of Wisconsin-Milwaukee, Center for Policy and Information Research: http://cipr.uwm.edu/?page_id=177

FURTHER NEEDS:

Addressing the privacy concerns that are relevant to our academic and research libraries that play a special role in guaranteeing intellectual freedom

Investigating and coming to consensus on the delicate balance between the need to collect and analyze user data in order to provide relevant services and the need to protect privacy

USER DATA WITHIN THE
SOCIOTECHINCAL SYSTEM OF INFORMATION

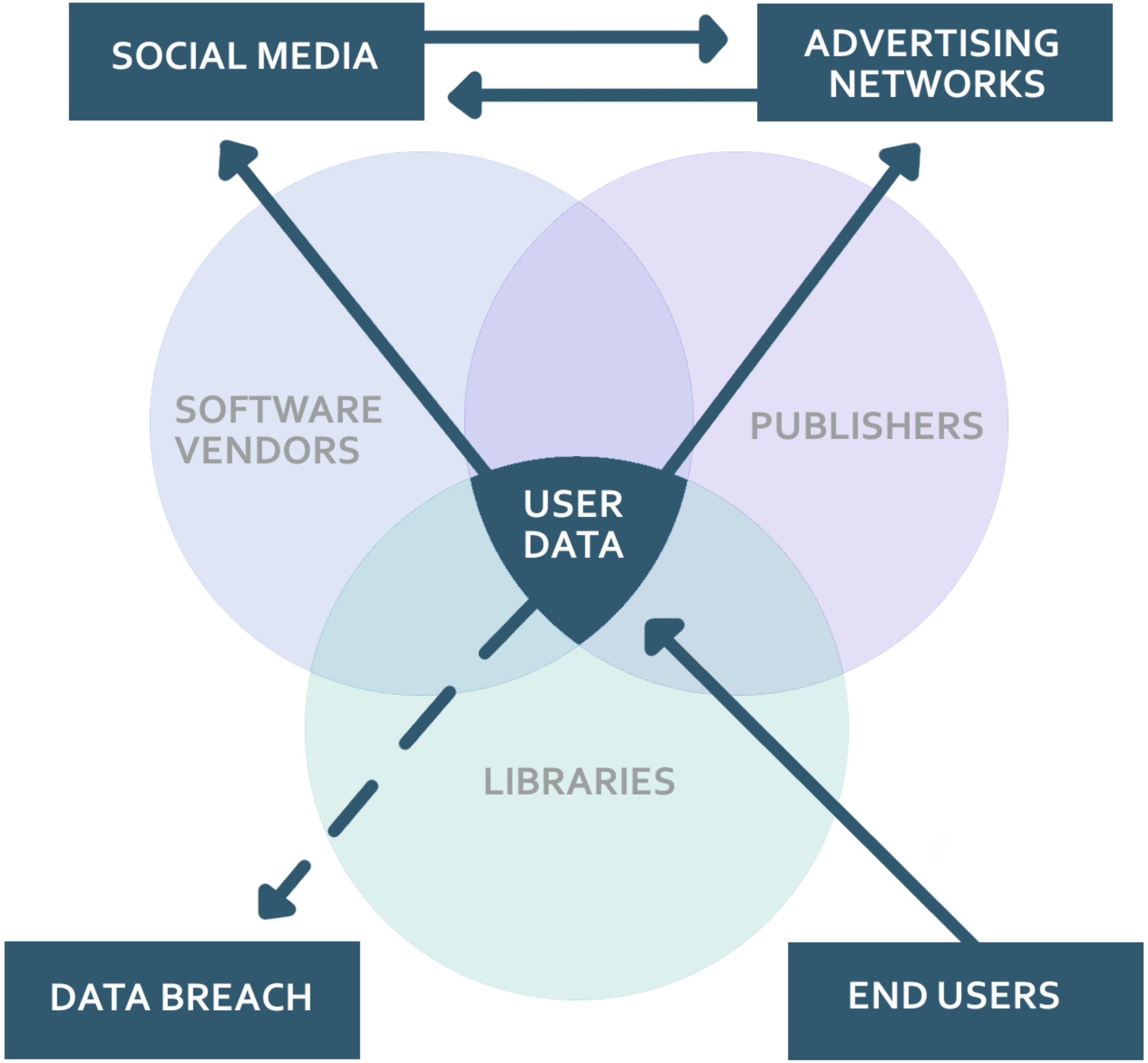


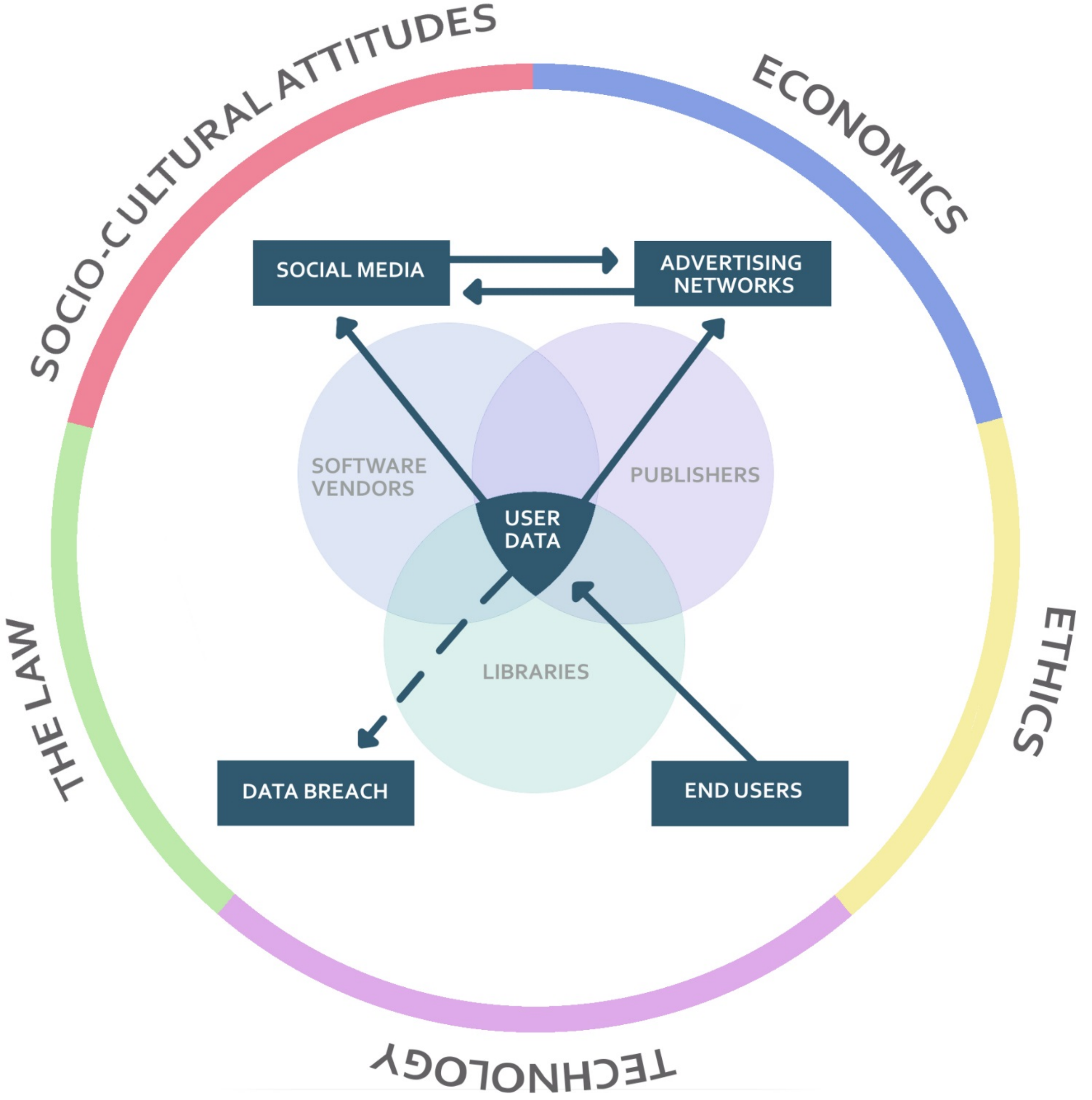
SOFTWARE VENDORS

PUBLISHERS

USER DATA

LIBRARIES


SOCIAL MEDIA
ADVERTISING NETWORKS
SOFTWARE VENDORS
PUBLISHERS
USER DATA
LIBRARIES
DATA BREACH
END USERS


SOCIO-CULTURAL ATTITUDES
ECONOMICS
ETHICS
TECHNOLOGY
THE LAW
SOCIAL MEDIA
ADVERTISING NETWORKS
SOFTWARE VENDORS
PUBLISHERS
USER DATA
LIBRARIES
DATA BREACH
END USERS

MELLON GRANT TO THE NATIONAL INFORMATION STANDARDS ORGANIZATION

“Patron Privacy in Digital Library and Information Systems”

- Led by expert steering committee.
- Is holding a series of meetings involving librarians, publishers, software developers and vendors.
- Will develop a framework of principles on the privacy of user data for use by libraries, publishers and vendors.
- Considering library-hosted systems; cloud-based library discovery, ILS and others applications; publisher and aggregator systems; and the legal aspects of data sharing and data breach reporting.

OUTCOMES

- Framework could become NISO Recommended Practice, if approved by NISO members.
- Further grants, on a highly selective basis, to support adoption of framework in the not-for-profit community.
- A starting point for international engagement among US-based and non-US based publishers, libraries, and vendors.
- Framework will assist the Foundation and other funders in asking informed questions of our grantees.