Online Supplemental Material for

Big Data at The Big House:
What Do University of Michigan Patrons
Think About Their Library Data?

Delaney Jorgensen, Zoe Garden, and Craig Smith

Section 1: Interview and Focus Group Protocol

Intro Steps for Interviews

- Intro to interviewers (and roles for this particular interview -- lead vs. note taker)
- Description of study/interview process
  - The study is exploring U-M students' thoughts and opinions on the data the library collects about its users, including students.
  - The interview will last no longer than one hour, and you can ask to stop at any time.
  - You will receive a $25 gift card for your participation.
  - Participation in this study does not involve notable risks to participants.
  - There are no right or wrong answers to any interview question. We're just interested in what you think.
  - Results of the study may be published, but no information would be included that would identify you.
- Consent for study
  - If they've read the form ahead of time, just see if they have questions. Then ask if they consent to participate.
- Consent to record
  - Ask if they are okay with recording the Zoom session (and note that recordings will be deleted after the project is over).
  - (Note that it's okay if they don't want it recorded.)
- Any questions before we get started?

Intro Steps for Focus Groups

- Z.G. and D.J. introduce themselves and their role at the library.
- Call on each participant to introduce themselves and respond to the following questions:
  - What are your name and pronouns?
  - [For grad students] What are you studying at U of M?
  - [For faculty] What department do you work in at the university?
  - What would you say is the main way you use the U-M Library?

- Intro to interviewers (and roles for this particular interview -- lead vs. note taker)
- Description of study/focus group process
  - The study is exploring U-M students' and faculty's thoughts and opinions on the data the library collects about its users.
  - The focus group will last no longer than one hour, and you can ask to stop at any time.
  - You will receive a $30 gift card for your participation.
  - Participation in this study does not involve notable risks to participants.
  - There are no right or wrong answers to any question. We're just interested in what you think.
  - Results of the study may be published, but no information would be included that would identify you.
- Consent for study
  - See if there are any questions.
  - Then ask if they consent to participate.
- Consent to record
  - If someone indicated on the Qualtrics form that they don't want the interview recorded, do not record the interview.
  - If all participants indicated that they are ok with recording, acknowledge that and then start the recording.
  - ALSO -- note that recording is only used to flesh out notes and then be deleted.
- Set a few key ground rules for the discussion
  - We will monitor raised hands throughout the discussion, so please use that function when appropriate.
  - If you have a comment to share, please communicate verbally with the group. If you are having connection difficulties or come up with an answer to a previous question later in the discussion, you can also share ideas in the chat.
  - You can use direct chat with Zoe or Delaney if you are having technical issues or would like to leave the focus group at any time.
  - Unless a participant says something that's combative or offensive, we want to emphasize that all opinions are valid and should be respected during our conversations.
  - We want all group members to participate equally in this discussion; if you know that you have been participating a lot, you can step back and give others the chance to share. We also encourage quieter participants to speak up because we value your opinion. Zoe and Delaney will help facilitate the conversation more directly if there is an imbalance.
  - Everything said in this discussion is private, so we encourage you to keep people's identities and names confidential after this focus group session is complete. We will do the same in our data collection and analysis.
  - Ask if there are any rules others would like to add.
- Any questions before we get started?

Opening Questions

1. What year of your studies are you in at U-M?

2. What are you studying here at U-M? [FOR FRESHMEN AND SOPH] Or if you aren't sure, what are you interested in?

3. In what ways do you currently use the U-M Library? -- such as getting books, accessing full-text articles, using study spaces, meeting with library staff?
   a. [IF NOT FRESHMAN] How about before the pandemic?

Understanding of Library Data

Intro script:
When you use internet services like Facebook, Instagram, and Google, these companies collect data about you. This could be personal information like demographics, or it could be information about your behavior on the service; what types of posts you interact with, what you search, etc. Other institutions collect data too. For example, the university collects personal data on things like your family's financial background, or your demographics. They can also use digital tracing to track your location around campus; for instance, when you use your MCard to access a building.

4. When you interact with the Library's spaces and resources, what kind of data do you anticipate the Library is collecting about you?
   a. For book checkouts?
   b. When accessing online articles?
   c. When accessing Library spaces?
   d. When meeting with a library staff member?
   e. Other situations?

Thoughts/Feelings about Library Data

Intro script: On this website, which is part of U-M's Information and Technology Services, you can see some examples of what data the library collects on you. [go through the website].

5. What are your reactions to hearing about the types of data the library collects on you?

6. We're interested in your level of trust in the library, with regard to responsible data management? For example, some types of data mismanagement include data leaks, data selling, and the sharing of identifiable information. Do you have concerns about any of these, with regard to Library data?

7. Do you have ideas about how the Library could use the data it stores about students to do something useful for students?
   a. To improve academic experiences?

b. To improve the campus experience, for example connecting you to events or programs?

Thoughts/Feelings about Library Analytics

Intro script: Analytics involves collecting data about individual users and their practices to make inferences about users' behavior. Companies sometimes use those inferences and predictions to suggest products or target advertising. As we noted, you also generate data when you interact with the university and Library websites, buildings, and services. The library could also use the data you generate to engage in analytics. For example, the library could look at the connection between patterns of library use among students and its correlation to student GPA. Or the library could use patterns of book checkout behavior to generate recommendations for future checkouts.

INTERVIEWS

8. The library is considering using an analytics approach to data. How do you feel about this?
    a. [IF NO CONCERNS EXPRESSED]: Do you have any concerns about the library using the data for analytics projects?

FOCUS GROUPS

8a. The library is considering using an analytics approach to data. Do you see the ways that this approach to collecting data could be positive for library patrons?

8b. Do you have any concerns about the library using data for analytics projects?

Thoughts about Transparency and Protection

9. Now that you know a bit more about the types of data collected by the Library, are there certain types of library data that you think the Library should protect anyone from accessing (including people doing assessment and research)?

10. Do you think the library has any responsibility to actively reach out to you to inform you about the data they collect and store about you?
    a. [IF YES] How would you want this communication to take place?
    b. [IF NO] Why not?

(INTERVIEWS ONLY)
11. Social media sites and other online platforms often have "terms of service." By agreeing to these terms, users give these companies authorization to collect and possibly store data about their behavior on those sites. The library currently does not ask for students' authorization to collect their data. How do you feel about this practice?

Closing Questions

12. Are there things about the library and your data that you want to share -- things we didn't cover so far?

13. Do you have any other questions, ideas, or comments?

(INTERVIEWS ONLY)
14. What is the best mailing address for us to send you your gift card?

Section 2: Coding Schemes for Interview Questions

Q4: When you interact with the Library's spaces and resources, what kind of data do you anticipate the Library is collecting about you?

Coding categories used:
- Book
- Online resources
- Library spaces
- Staff member meeting
- Accessing databases
- Demographics
- Other situations

Q5: What are your reactions to hearing about the types of data the library collects on you?

Coding categories used:
- Not surprised / expected
- Explicit expression of not being concerned
- Surprised at amount of data collected
- Some and/or all of it feels invasive
- Some of it feels unnecessary
- Trusts that the library is only collected what it needs to serve patrons
- Concerned / surprised about data retention
- Concern with lack of notification and/or consent
- Understands why some people may have concerns
- Concern about data shared with / held by third parties
- Concern about use of Google analytics
- Uncodable / difficult to code

Q6: We're interested in your level of trust in the library, with regard to responsible data management? For example, some types of data mismanagement include data leaks, data selling, and the sharing of identifiable information. Do you have concerns about any of these, with regard to Library data?

Coding categories used:
- No concerns at all
- Mildly concerned
- Resigned to data collection
- Concerned after learning new info about library data
- Trusts library to be responsible with data
- Trusts the library more than the university as a whole
- Trust in the library is equal to trust in the university

Q7: Do you have ideas about how the Library could use the data it stores about students to do something useful for students?

Coding categories used:
- Suggest materials
- Filter databases
- Expand library catalog
- Personalize communication with students
- Suggest events
- Create events around frequent student questions or student interests
- Increase access to popular materials (i.e. databases, subscriptions)
- Create aids or learning initiatives to help student learning
- Connect students to summer opportunities
- No ideas

Q8: The library is considering using an analytics approach to data. How do you feel about this?

Coding categories used:
- In favor of using analytics on student data
- Does not think they would be personally affected by analytics
- Uncertain / needs more information
- Not sure if student data is applicable to library analytics
- Concerned about data privacy
- Concerned about identifiable information
- Concerned about bias in technology for social identities
- Concerned that data collection will replace real conversations with students
- No personal concerns; understands that other students may have data privacy concerns
- No pressing concerns

Q9: Now that you know a bit more about the types of data collected by the Library, are there certain types of library data that you think the Library should protect anyone from accessing (including people doing assessment and research)?

Coding categories used:
- Identifiable information should be protected
- Information should not be shared with third parties outside the university (including local, state, and federal governments)
- Location information should be protected
- Sensitive demographic information should be protected
- Financial information should be protected
- General online activity should be protected
- OK with the library sharing identifiable data
- OK with the library sharing demographic data
- Not concerned about data sharing
- Understands why other students may be concerned about data sharing
- No ideas

Q10: Do you think the library has any responsibility to actively reach out to you to inform you about the data they collect and store about you?

Coding categories used:
- Yes, students should be informed of all data collection by the library
- Yes, students should be informed when their personal data changes
- Yes, but only to summarize what's being collected periodically
- Yes, but only in the case of a data breach or leak
- The library should provide information, but does not need to actively reach out to students
- Students would not be interested in learning about their data
- The library has a right to collect data without informing students
- Favors opt-in / opt-out system of collection

Q11: The library currently does not ask for students' authorization to collect their data. How do you feel about this practice?

Coding categories used:
- Yes, the library should ask for student authorization
- Yes, but no in-depth information is necessary
- No personal concerns
- Understands why other students may be concerned
- Authorization would not change how students interact with the library
- Authorization makes students pessimistic about data collection
- Expects data collection from the library
- Was unaware that the library did not ask for authorization

Concrete suggestions for the library to implement an authorization system (# of undergraduates who mentioned each suggestion):
- Terms of service (3)
- Keep terms of service shorter than one page (1)
- Minimal communication (3)
- Model Wolverine Access' consent process (2)
- Opt-in/opt-out, and discuss with students (1)

Section 3: Coding Schemes for Focus Group Questions

Q4: When you interact with the Library's spaces and resources, what kind of data do you anticipate the Library is collecting about you?

Coding categories used:
- Information on materials accessed
- Patron data
- Session information (i.e. length of session, downloads)
- Data for recommendations
- Did not know / unsure if library collects data

Q5: What are your reactions to hearing about the types of data the library collects on you?

Coding categories used:
- Surprised
- Not surprised
- Has questions
- Concerns about data collected
- Collecting more data than expected

Q6: We're interested in your level of trust in the library, with regard to responsible data management? For example, some types of data mismanagement include data leaks, data selling, and the sharing of identifiable information. Do you have concerns about any of these, with regard to Library data?

Coding categories used:
- Trusts library to be responsible
- Trust in university as a whole
- Institutional framework (i.e. laws, IRB) protects data
- Data are not interesting / vulnerable enough
- Concerned in some way

Q7: Do you have ideas about how the Library could use the data it stores about students to do something useful for students?

Coding categories used:
- Recommendations
- Identifying areas for further research / subfields
- Improving databases
- Troubleshooting online resources

Q8a. The library is considering using an analytics approach to data. Do you see the ways that this approach to collecting data could be positive for library patrons?

Coding categories used:
- Benefits for research
- Benefits to connect to other communities
- Beneficial for library resources / improving resources
- Beneficial for the library as a whole
- Beneficial for patrons / user experience

Q8b. Do you have any concerns about the library using data for analytics projects?

Coding categories used:
- Recommendation concerns
- Opt-in / opt-out concerns
- Overall data concerns
- Patron data concerns
- Diversity concerns
- Scholarly / academic concerns
- Library overstepping boundaries

Q9: Now that you know a bit more about the types of data collected by the Library, are there certain types of library data that you think the Library should protect anyone from accessing (including people doing assessment and research)?

Coding categories used:
- Protect library materials
- Protect patron information
- Protect login information
- Protect data from outside access
- No concerns

Q10: Do you think the library has any responsibility to actively reach out to you to inform you about the data they collect and store about you?

Coding categories used:
- Reach out regarding patron data
- Opt-in / opt-out
- Email message
- Library is not responsible for communication
- Library is responsible for communication
- Unsure about responsibility of library
- Patrons assume data collection by default
- Reach out once a year (offer an orientation, or send a one-time message)