IMT 570 B:

ANALYTIC METHODS FOR INFORMATION PROFESSIONALS

Group Assignment 2: Research Design & Method

*“Is there a need of an open data repository for academic purposes at the iSchool, UW?”*

Submitted by-

Group 7:

Yanjie Niu

Sejal Khatri

Ishita Bhandari

Prakirn Kumar

Abstract

Open source data is the idea that some data should be freely available to everyone to use and re-publish as they wish, without restrictions from copyright, patents or other mechanisms of control (Auer et al., 2007). An open data repository for academic purposes will allow instructors, researchers and students to collaborate and access the data like, but not limited to, research materials, dissertations or teaching material.

Currently, the Information School at the University of Washington doesn’t provide a unified open data portal for its community to store and share their data on a single platform. As a researcher if you want to know about the research being done in the university and want its access then you have to jump through a  few hoops and try to find relevant materials posted on different portals. This is exactly what sparked a series of questions in our mind - What are the expectations and major concerns associated with researchers when it comes to sharing their data on these platforms. With schools like Harvard, Georgia Tech, etc. taking part in the open data movement, whether or not, is there a need for such a repository at the iSchool at UW.

Aim

We aim to analyze the need of an open data repository which serves as a unifying platform for publishing and sharing data related to the research and other academic activities going on in the iSchool, UW. The emphasis is on unearthing the underlying reasoning behind people’s response to the need of an open data repository.

Objectives

* To identify if there’s a need of an open data repository at the iSchool.
* To explore the reasons behind people’s responses.
* To obtain exact needs and expectations of the people regarding the topic.
* To check the feasibility of creating a repository to meet those needs.
* To provide recommendations to the iSchool.

Methods for Data Collection

We are following the mixed approach for data collection, taking advantage of both Qualitative and Quantitative research methods, as this will help us in obtaining data that can be measured and analyzed and data that cannot be measured. Our primary data collection method are interviews and surveys. This will help us understand the current scenario of the open data availability and usage at the iSchool and the opinion and needs of several professors and students who are involved with such data on daily basis. For secondary data collection methods we will integrate existing data available online or data collected during similar types of researches. We are planning to evaluate the credibility of qualitative data by using the quantitative data.

*Sampling methods:*

1. *Snowball Sampling*   
   Snowball Sampling method is the method in which participants refer the researcher to others who may be able to potentially contribute or participate in the study. Since we are unaware of the research interests of the professors, we can use this method to help professors introduce us to researchers who are working in the same area.
2. *Purposeful Sampling*  
   It is the method in which participants are selected or sought after based on pre-selected criteria based on the research question. Since, our research problem consists of the term ‘Open Data’, we feel not everyone will be able to understand it properly. Therefore, we have decided to use purposeful sampling to target PhD students and professors working in the field of data.
3. *Quota Sampling*  
   It is the method in which we would gather data from a certain number of participants that meet certain characteristics.

*Qualitative Data Collection*

Our selected participants for Interviews are iSchool professors who we see as the potential contributors and users of open data repository along with some professors who might have different opinions to get a holistic view. We are also planning to interview PhD  students who are doing research in the field of open data and taking initiatives for open data movement. PhD students will help us understand the complexities of our research and can be of great resource for directing us to professors with similar interests. We plan to interview 5-8 professors and 3-5 PhD students for our research.

*Example of some of the questions we’ll ask:*

1. Think about the recent research you have done. Would you publish the data you collected during that research on the open data portal provided by the iSchool?
2. If you'll publish data how much would you publish on a weekly basis?
3. What are your concerns about the open data portal?
4. How do you think will students benefit from this?
5. What features would you want in the repository?
6. Who do you think would benefit from open data repository?
7. Who do you think will be able to help with this research?
8. Would this research be useful? (Would you use the open data repository? How?)

*Quantitative Data Collection*

The interviews conducted by us, of various professors and students will help us in collecting a vast array of qualitative data, but in order to check the credibility of the data and insights that we obtain from qualitative data, we need to resort to quantitative data.

For the quantitative data collection, we will send out a survey to professors at the iSchool, who have a lot of data related to their course instructions and the researches they work on. We will also survey PhD students, who are involved in various researches and Master’s students who who might be interested in reading about ongoing researches in the iSchool. We plan to send out about 100 surveys for the professors and students combined and expect to have around 30% response rate.

*Through our survey, we would like to get and analyze answers for several questions, few of which are listed below:*

1. Are you interested and excited about the idea of iSchool having an open data repository?
2. Are you willing to share your course instructions and resources on an open data repository of the iSchool? (For the professors only)
3. Are you willing to share your research paper or projects on an open data repository of the iSchool?
4. On scale of 5 (5 being most useful), how useful it is having an open data repository consisting of iSchool professors’ course instructions and research data of iSchool professors and PhD students?

Plan of Work

1. *Discover, Identify, Define, and Refine the Research Problem (Oct. 1 - Nov. 1, 2018):*

* Through discussing and interviewing researchers and professors at iSchool, UW, in October, we identified the needs and concerns toward the issue of open data repository.
* Then, we went through the existing information and data to do a literature review and understand the status quo of open data repository in contemporary higher education. We got to know different universities and disciplines have different policies and practices regarding open data repository. iSchool and other peer departments at UW do not have its individual open data repository.
* Considering the 10-week length, we scoped down the research question to the need of open data repository for iSchool at UW.

1. *Research Proposal (Oct. 5 - Nov. 1, 2018):*

We finished our first research proposal, got feedback from Meg, Richard, and Nami. Based on the feedback, we went back to the field and literature to continue refining our research plan and topic.

1. *Research Planning (Oct. 10 - Nov. 1, 2018):*

* We will use both qualitative and quantitative research design strategies.
* *Sampling Design:*

We want to make the sample more representative, so it would include diverse groups of people in iSchool.

Researchers: For example, data scientists in the data lab, who care about the data usage and sharing a lot in their quantitative research.

Faculty Members: For example, teaching faculty members who pay attention to user datasets in instruction a lot.

Nicholas Weber and Meg Young would help us connect those researchers. This would be done 11/1 - 11/4 to contact them.

* *Quantitative Data Collection Design:* Sejal leads this and we have improved it through group discussion (11/1).
* *Qualitative Data Collection Design:* Prakirn leads this and we have improved it through group discussion (11/1).
* *Pilot Testing (11/1- 11/ 16):* Yanjie and Ishita would test the questions for both survey and interview. Our consideration would be about ethical management, suitability, and feasibility. We would also keep adjusting the questions, if in the field when doing surveys and interviews we find there are some things we need to address. So some pilot testing would start earlier than the formal ones.

1. *Implementation:*

* Gather and Analyze Quantitative Data:

First is through sending surveys (11/2 - 11/5). We have designed surveys and need to send surveys to teaching faculties, researchers, and PhD students in School.

Ishita will email Meg and Nic for connecting respondents. Then, our team send surveys to them. We aim for a sample of 30 participants and understand the response rate might not be high like other surveys. So we would send out around 100 surveys if possible to expect at least 30% respond.

Then, we wait for surveys return (11/5 - 11/9).

While receiving returned survey results gradually, we start to use excel to do

quantitative data analysis (Prakirn and Ishita leading, 11/5 - 11/9).

* Gather and analyze qualitative data.

Based on preliminary survey results, we started to adjust interview questions and do interviews to collect qualitative data (11/9 - 11/16).

We would first identify and reach out to 6-10 respondents and schedule appointments with them (11/5 - 11/9).

Afterwards, four members are grouped into two teams and pair up to do interviews (11/9 -11/16), in which one person asks questions, while the other takes notes. Yanjie and Ishita as one team, Prakirn and Sejal as the other team.

Then, we would do analysis, based on notes and observation in the interviews.

1. *Report Writing and Slides Preparation:*  
   Based on data analysis, we work on report.

11/1 - 11/5, introduction and literature review.

11/6 - 11/16, start to write the outline of main body, based on data we collect.

11/17 - 11/21, write main body and conclusion, and make visualized data for presentation.

11/22 - 11/ 25, go through report draft and send it to professors and TAs for feedback. Then, start to make slides.

1. *Research Reporting and Presentation:*

* *Presenting Practice: Nov 26 - 29*  
  We can practice with each other in the beginning three days of the week before mock presentation in class.
* *Formal Presentation: Dec 4/6*  
  Improve presentation based on practice and advice from professors, TAs, and guests.
* *Final Research Report needs to be done on or before Dec 4*  
  Need to finish the final research report and polish it based on the feedback from professors, TAs, and cohorts in the Q&A session.

One benchmark is to make findings and recommendations in our report evidence-based, comprehensively analyzed, and actionable.

Risks and Ethical Considerations

There are many risks associated when it comes to research on open data. Open data portals have been continuously surrounded by controversial opinions and we have identified the following major risks and ethical concerns associated with our research:

* Since the aim of this research is not to pick any side of the argument - “Whether or not does iSchool need an open data repository”- but to identify the current scenario and present the views as they are, we will have to be very careful of not asking any leading questions in our survey and interviews.
* Many a times, researchers unknowingly influence the research with their opinions and this may result in a manipulated research. This especially can be a challenging task as all four of us feel passionately about this topic and hold certain views which, if kept unchecked, may result in inaccurate analysis.
* Getting consent from the participants for interviewing them and using their views in our research is very important especially because we might be dealing with information which may seem harmless on the outside but can be sensitive.
* Maintaining anonymity of the responses and protecting our participants from any kind of harm will be our priority and a huge challenge. The data collected may be sensitive and thus, has to be protected and anonymized effectively.
* Researchers, for a couple of reasons, might not be very honest in their response as to why they would or wouldn’t want to publish their data.
* Since a huge portion of our participants are being chosen through snowball sampling hence, there are chances that we might focus too much on certain populations or miss other population altogether.

References

* Open Knowledge International. What is Open Data? Retrieved from: <http://opendatahandbook.org/guide/en/what-is-open-data/>
* Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. The Semantic Web. Lecture Notes in Computer Science, 4825, 722.