Xinyun Zhang
xinyunzh@umd.edu

Summary:

My client is Environmental Finance Center(EFC) in the University of Maryland, who is dealing with reports data generated by organizations using money from 9 funders. EFC collate Water Quality and Resiliency Financing information, capturing the inputs and corresponding outputs of projects across the region. What they need to do is to rate whether a project is successful or not and write the final analysis report to let CBPO know how to better invest their money and which organizations can be potentially more successful.

One thing surprised me is the database structure. When talking with the database designer, I was surprised that funders can also be implementers, meaning it will be way more difficult for database designer to structure the relationship between two entities.

Detailed interview memos will be in the appendix.

Key challenges:

1. Too much data and too less data
    - Too much data for EFC staff to manually go through and data keeps growing.
    - From the database designer's perspective, he needs more data to validate the structure of the database. Meanwhile, EFC staff need more data to prove the reports they have. For example, they want to conduct follow-up interviews with people who participated in implementers' projects to see if projects really work well.
2. Database building problem
    - Samarjith is the only guy who is building the database and he is also the only one who has technical skills in EFC. He said he might need one more person to discuss as it is hard for him to design the database himself.
    - Data keeps growing and he needs to keep revising his database structure.

Appendix

The 1st interview

Researcher info:

Name: Brandy Alexis Espinola

Email: bespinol@umd.edu

Role: Project manager who is in charge of Track 3 project

School: The University of Maryland

Project Info:

Project title: Informing, improving, and expanding Water Quality and Resiliency Financing through Advanced information management

Project duration: April 2017- August 2017

Project context: Collate water quality and resiliency financing information capturing the inputs and corresponding outputs of projects across the region. Our work will initially focus on capturing information related to projects funded directly by EPA and the Chesapeake Bay Program, as well as projects the EFC has conducted over its nearly 25-year history.

Defining your data:

1. Data Description:

Both quantitative and qualitative data with regard to projects funded by EPA with Chesapeake Bay projects.

2. Where does your data come from?

Organizations(implementers) will hand in reports to funders of Chesapeake Bay region. Funders gather all reports given by organizations using their money and put reports into a csv file and send to EFC.

3. How much data do you generate?

We don't generate data. What we want to do is to put reports data into a database, so that we can analyze easily.

4. What file formats do you use?

Mostly text file. For example, csv, pdf. Information stored in the database is also text format.

Looking after your data:

5. What different versions of each data file do you create?

The only thing we generate is database to store data.

6. What metadata are you adding to each data file?

We haven't added any metadata into data file.

7. Where do you store your data?

I store my data in the cloud and save locally in my computer. I don't worry about the hardcopy because reports generators have the original hardcopy.

8. How do you structure and name your folders and files?

I didn't have a plan on that but I will probably rename my files.

9. How is your data backed up?

I haven't backed up my data but I plan to.

10. How will you test whether you can restore from your backups?

Since I haven't backed up yet, I haven't tried testing.

11. Who is responsible for the immediate day-to-day management, storage and backup of your data?

Me

Sharing your data:

12. Who owns the data you generate?

Funders

13. Are you working with collaborators (sharing your data with them)?

Yes, we do. We now have 5 persons (Dr. Sprinkle, me, database designer, an intern reviewing reports, and you) working on this project. Reports providers (9 funders) will not be able to access our database. However, they can see the final reports. The way to exchange information is via email.

14. What should and shouldn't be shared and why?

Nothing cannot be shared within 5 people, but we can't provide information to other people, for example, the funders.

Archiving your data:

15. What should be archived beyond the end of the project

I don't know what should be archived beyond the end of the project but I definitely plan to archive information afterwards because it should be a long-term project.

16. For how long should it be stored?

Probably 5 years.

17. When will files be moved into the archive?

After we get the final reports.

18. Where will the archive be stored?

I don't know, probably on the cloud.

19. Who is responsible for moving data to the archive and maintaining it?

I guess me.

20. Who should have access and under what conditions?

Funders and implementers can see final reports but they can't access the database.

Data management plan questions:

21. Who is responsible for making sure this plan is followed?

Me

22. How often will this plan be reviewed and updated?

3 years? I'm not sure.

23. What actions have you identified from the rest of the plan?

Since we haven't finished the plan yet, I don't know what specific action we will take. We probably will follow what the plan says. Say, ask someone to analyze data for us.

24. What further information do you need to carry out these actions?

We need to build the database first and review all the reports.

Data services questions:

25. What data services would assist you in your research?

I don't know that, probably analyzing.

The 2nd Interview:

Researcher info:

Name: Samarjith Jawaharlal Sathyanarayan

Email: samarjit@terpmail.umd.edu

Role: database designer

School: information school in the University of Maryland.

Project Info:

Project title: Informing, improving, and expanding Water Quality and Resiliency Financing through Advanced information management

Project duration: April 2017- August 2017

Project context: Collate Water Quality and Resiliency Financing information, capturing the inputs and corresponding outputs of projects across the region. Our work will initially focus on capturing information related to projects funded directly by EPA and the Chesapeake Bay Program, as well as projects the EFC has conducted over its nearly 25-year history.

Defining your data:

1. Data Description.

Quantitative and qualitative data.

2. Where does your data come from?

National Fish and Wildlife Foundation-NFWF

3. How much data do you generate

Filemaker pro will store data automatically. Data will be put on the server. However, I will manually copy each thing.

4. What file formats do you use?

Csv, pdf

Looking after your data:

5. What different versions of each data file do you create?

Database

6. What metadata are you adding to each data file?

After building the database structure, I will start to document, taking notes on Filemaker pro.

7. Where do you store your data?

Only local right now.

8. How do you structure and name your folders and files?

Name each columns of database.

9. How is your data backed up?

I make copies locally. Before every check, I will back up.

10. How will you test whether you can restore from your backups?

Make copies to ensure I can go back when I might go wrong in database designing process.

11. Who is responsible for the immediate day-to-day management, storage and backup of your data?

Maybe NFWF, and I don't know how owners will backup.

Sharing your data:

12. Who owns the data you generate?

NFWF

13. Are you working with collaborators (sharing your data with them)?

Yes, 5 people.

14. What should and shouldn't be shared and why?

We don't have things we can't share within us 5 people.

Archiving your data:

15. What should be archived beyond the end of the project?

I don't know things about archiving but I guess we will.

Since Samarjith is only responsible for the database design, we didn't cover this part.

16. For how long should it be stored?
17. When will files be moved into the archive?
18. Where will the archive be stored?
19. Who is responsible for moving data to the archive and maintaining it?
20. Who should have access and under what conditions?

Data management plan questions:

21. Who is responsible for making sure this plan is followed?

Brandy

22. How often will this plan be reviewed and updated?

I will review database every week but not necessarily update

23. What actions have you identified from the rest of the plan?

I don't know. I will continue designing the database.

24. What further information do you need to carry out these actions?

I probably need one more person to help me design the database. I want someone I can talk with so that I might get inspired by him/her.

Data services questions:

25. What data services would assist you in your research?

Probably one more person to help me.