Data: What is it?

Reading reflections - Simon Chen

The readings in this week share ample interesting data application examples in different cases and brought me a lot of challenging problems which I never thought about before. They provide perspectives of the nature of data and some key aspects of their life-cycle in the context of data reu-use (Carlson & Anderson, 2007), from research data to educational data which covered almost every aspect in real life. The articles provide some cases of data collection, data formatting, data release and data re-use, with more practical challenges examples. And one of the articles points out the importance of analyzing softwares which had a profound relationship between research sustainability and reproducibility (Jay et al., 2021). Also, there is one more detail about the research data, which includes types of research data, data life cycle, data management practices (storage, backups, version control, data security and data curation and preservation) with several case studies (Scott et al., 2016). All of the papers I mentioned above could contribute to the last reading - Toward the Relational Management of Educational Measurement Data (Chung, 2014). The data management and analysis in education could help instructors to understand the student, including static data, low-frequency data and high-frequency data, because the era of big data also merged into the educational industries with transmedia-based learning and personalized instruction.

As I said above, there are many practical challenges in real life. For example, in the AnthroProject, it would take the project 30 years to digitize their research materials progressively and 20 years for the CurationProject to document its entire collection of over 850,000 objects in its database. Also it mentioned that “the problem is that you might not know that it’s sensitive knowledge until it gets public…” (Carlson & Anderson, 2007). And in educational analysis, well-designed user interfaces may work against the measurement of user intent by those too-easy user interfaces. I just listed two examples here but I saw more problems under the surfaces.

With some challenges I mentioned above, it reminds me of my previous working experience in a company as a data engineer. Before managing the customer's database system, we have to successfully collect the customer data. Due to the certain situation (we provided some services to the aged people, and they can't use online tools to digitize their information), the customer's information was collected by the paper forms, so an initial process of translation of data into digital form is needed. But this process is time consuming, exhausting and might be a different and complementary kind of data. As the company's business grows fast, more customer information needs to be documented digitally. At the beginning, we spent months organizing and collecting the data. As in the reading mentioned, the task of digitizing data derived from other media proves very costly and sometimes requires decades in itself (Carlson & Anderson, 2007). Employees complained about the heavy duties, and the boss cried for the money wastes, then we made some changes to the data collection procedures. We provided more simple and specific online filling and register instructions, so that the data we collected could be directly loaded into our database. Since there are some customers who can not even read English, then we provided multiple language choices to help them understand the forms. Even though our company is not big, we still spent tons of time digitizing data to make it useful for market analysis. I can't imagine how much time and effort the government would need to do a society level data analysis. "At the present date only 30% of the collections in the U.K. are digitized." I can't agree more.

I also got inspiration from the Introducing research data. Before the data collection, you could categorize the data and think about what needs to be done with it. You should find out if your discipline already has standard practices and use them (Scott et al., 2016). Usually, we just used the data to do analysis, but seldom to think about the data collection. I rethink my personal working experiences and the real case studies in readings, I found some useful information to improve the data collection efficiency in future.

At the education level, the challenges in learning data collection make me think about my own online learning experiences. The use of Canvas provides us some convenient features but I am still wondering if the data collected by Canvas could bring useful information to instructors to provide more

personalized and effective learning experiences? How could we improve the online learning experience with more and more advanced learning data analysis? That is part of the reasons I choose to study in the learning analytics program.

References

Carlson, S., & Anderson, B. (2007). What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication*, *12*(2), 635–651. https://doi.org/10.1111/j.1083-6101.2007.00342.x

Chung, G. K. W. K. (2014). Toward the Relational Management of Educational Measurement Data. *Teachers College Record*, *116*(11). http://www.tcrecord.org.ezproxy.cul.columbia.edu/library/content.asp?contentid= 17650

Jay, C., Haines, R., & Katz, D. S. (2021). Software Must be Recognised as an Important Output of Scholarly Research. *International Journal of Digital Curation*, *16*(1), 6. https://doi.org/10.2218/ijdc.v16i1.745

Scott, M., Boardman, R. P., Reed, P. A. S., Byatt, D., Stark, I., Cox, S. J., Coles, S. J., Collins, A., Earl, G. P., Jasion, G., Mbuya, T. O., & Soady, K. A. (2016, October 31). *Introducing research data—Fourth edition* [Monograph]. University of Southampton. https://eprints.soton.ac.uk/403440/