

Saaif Ahmed

saaifza2

CS:598 Foundations of Data Curation

Assignment 3: Part 2

Dear Government Director,

Hello. I am Saaif Ahmed, a new hire for the department working in the data science department in the agency. As part of my current project, I have been given the task of overseeing the transfer of the old system to the new system. This involved managing the transition of the data not only from hardware location, but the transformation of the data to ensure the equivalence and accuracy of the data between the old and new systems. As direct overseer for this, it has been incredibly apparent how important curatorial work is for tasks such as these. Though, it has come to my attention that you have decided to adjust department funding over the next couple of months, and this adjustment will come at a major cost to my team and its work. I am writing this note to you to hopefully advocate for my team in the best manner possible and implore upon you the importance of Data Curation not only for the agency, but Data Science in general.

To begin we must first introduce Data Curation within the broader concept of data science. Data science is a well-known field for analysis of data. Data science uses various statistical, scientific, and computational methods to extract knowledge and insights from all forms of data be it pure, raw, structured, or unstructured. This is not the whole of the story, as what has been described can be defined as data analytics, as it is concerned with analytical techniques to extract knowledge from data. However, when it comes to data, and information in general, there must be a way to ensure that the data can be reliably and efficiently found and used. This is where Data Curation makes its entrance. Data curation is concerned with the management of data such that analysis later performed it better supported. For the purposes of our agency, this means that my team will make the work of the rest of the agency much easier as we develop standards for the complaints of customers. Then, when improvements are made to the system, much like they were during the project I was overseeing, the transition is much smoother and less costly otherwise. For the benefit of the agency, data curation is how the data the agency relies on is future-proofed from many errors down the line.

One of the objectives of data curation is managing the discoverability of the data the agency stores by wielding the metadata. As we are an agency managing complaints for financial institutions, it is not hard to imagine the possibility that some of the agencies we hold complaints for are abusing their customers and their policy. In such cases a class action lawsuit can be filed and for that class action, it falls upon our agency to provide everyone involved with the data necessary for the court proceedings. Without my team, the data would not be managed, and the discoverability of the data would completely disappear. We manage the metadata of every single complaint sent to us by customers within the country, which means we directly oversee how easily searchable, readable, and usable the data retrieved by any individual is. As time moves on and standards change, and the information the agency manages is added to, the metadata is changed to accommodate that. Thus, if the government itself wants to take a company to court to ensure justice for its citizens, our agency can provide the metadata in which they may search for whatever they may need because of my team curatorial activities.

The task that I am the overseer for is a prime example of Data Curation being necessary for data abstraction. Since you are a new director, you have not seen many changes throughout your career here at the agency. However, even though I am a new hire, I am familiar with the previous iterations of the data we manage here, as that is a curatorial activity. Director, I encourage you to think about how this transition from the old system to the new system is not the first time that the agency decided “out with the old in with the new”. Data Curation, and the work my team does are the reasons why we can make these changes quickly and efficiently. Right now, we take complaints, by mail, phone, and online form submission. Building 3 data systems for each method of submission prevents integration with each other, let alone further analysis down the road. The legal department cannot afford to sift through 3 sets of data, all different from each other, to determine how many infractions company X has caused over the past fiscal year. With my team, we can design schemas, structures, and forms of data management that abstract away from the physical representation of the data. With our designs we can preserve all the information we need to operate as the agency, while still opening our doors to everyone and getting an accurate landscape of financial institutions in the country right now.

These are some of the reasons why I would implore you to consider the Data Curation team for what it is to the agency a crucial asset. I am serious when I say that our team and Data Curation as a discipline is the reason why the agency has seen any success at all over the past few years. As the volume of data input has gone exponentially up the need for a time like mine has as well. Without it the delay times for our agency would be unimaginably worse and prevent real change from happening in the country. Please consider my department when building the budget for next year. Thank you.

Sincerely,

Saaif Ahmed