

To: Director
From: SWAPNIL DARMORA
Date: 29/11/2020
Re: Importance of data curation

Our organization deals with complaints made by different customers against various financial institutions as such we need to follow rigorous standards to maintain the integrity of the data as well as the migration of data from one system to another. Since we are a government organization, we are responsible for data of the customers and we need to support various workflows and ensure the authentication and authorization of data as well. Therefore, we need to understand the importance of data curation for our organization. We will address few touchpoints as to why this is important for our organization.

Data curation is active management of data through its life cycle of interest and usefulness. Some of the key areas of data curation are Collection, Organization, Storage, Preservation, Discoverability, Access, Workflow, Identification and Integration. Collection supports collection and acquisition of data from various sources and such is one of the important key area for our organization to focus on because we collect data from various customers about various financial institutions. Organization involves employing an appropriated data model and this is also required because we deal with vast information and in order to organize the data, we need to follow some data model so that we understand the underlying structure of the data. Preservation ensures that the data is understandable in future which is crucial to our organization as well when we migrate our data from one system to the another we need to have a broad understanding of the data present in the old system. Discoverability supports ability to search and locate for the data and this ensures that we will be able to search the complaints using their ids and locate the data related to them. Additionally, we need to ensure that the data that we have from customers can be distributed but only post identification and authentication, this is when access, identification and workflow come into picture. Integration supports data from different sources and it applies to us as well since we collect complaints from web, phone and other sources.

I would like to introduce you to the concept of meta data. Metadata is essentially the data about data. Meta data is very crucial for long term management of data. Metadata documents how the data is represented and how and where the data is copied. This helps in integration of data from one system to another as we have done recently from old system to new system. Metadata is very crucial for reformatting our existing standards as we need to understand the current data that is documented. Metadata also ensures that we have proper access and distribution of data.

Since we are a government organization, we need to have data polices and regulations in place because we need to maintain the privacy of customer complaints and data that we have. We have sensitive information like credit card details, contact details and financial details about our customer and we need to follow policies so that we comply with security, privacy rules. One of the challenges here is that we get complaint from various locations and hence we need to have sharing polices as well as policies for responsible conduct. We need to address how we store our data, how do we redistribute our data and how do we plan to archive data. We need to define who will have access to what data.

Since we are a government organization we are bound to keep this data safe and usable for the future and hence we also need to focus on preservation. Preservation ensures that we ensure reliable communication with the future. Since we will have a lot of researches that will come in possession of the data we have on different systems we need to ensure that they correctly recognize the data that we have present in our systems. Preservation helps us in that. It will ensure that data can be correctly read and understood by someone who will be looking at it in the future. This can be done by replication the data , migrating data to new formats because we will have to keep up with the trends and at the same time maintain the intended context of the data. We also need to convert data to standard formats optimized for preservation so that using standard methods we can correctly identify the data as it is in the future. We will b creating new data or developing insights from the data about the financial institutions that we have but we should also be able to retrack the source of the data for those insights. This is where data provenance comes into play, we should have a chain of information where source of data can be tracked.

Since we are a government organization we need to follow these curational activities to ensure that we have maintain privacy and integrity of data present in our systems