Write a convincing memo (650 word max) explaining why data curation services are important. Assume that the memo is written for your new director, who is not familiar with data curation, and not convinced whether to keep funding this work. You will want to make sure to introduce data curation within the broader context of data science. You will need to cover the key areas that you think are the most important for data curation at your company. We ask that you incorporate at least two of the following topics into your memo: Provenance, Policy, Metadata, and/or Preservation.

**Memorandum Seeking Support for the Data Curation Activities Associated with Government Agency XYZ’s Migration to the New Customer Complaint System**

**MEMORANDUM**

**DATE**: 11/12/2018

**TO:**

The Director of Data Science,

Government Agency XYZ .

**CC:**

Data Science Department

**FROM**:

Data Scientist ,

Government Agency XYZ .

**SUBJECT**: Request seeking support and funding from Director of Data Science for the data curation activities related to customer complaint system migration project

Sir,

We have come to know of a decision from Government Agency XYZ’s Data Science department to de-prioritize the data curation activities that are being planned as part of data migration to the Agency’s new customer complaint platform. This memorandum requests you to re-consider that decision and provide continued funding and support for the data curation activities to ensure successful data migration and business transformation.

I would like to present some findings and resulting thoughts from my latest assignment which was to assess the quality of the recent data transfer from our complaint system to the new one, in support of my request to re-consider the decision. Below are some of the key points to which I would like to draw your attention to.

As we migrate our data sets from our old system to the new one, it is very important that we avoid any possible cases of potential data loss. But assessment of sample data from both the systems show that data in the old and new systems are not 100% the same. We all know that preservation of data – be it historical or current – is very important for the successful day-to-day functioning of our agency. Any data loss or incorrect data in the new system is going to affect the agency’s ability to support successful and fast-paced customer complaint resolution as expected by the involved parties. Missing data will create gaps that might lead to unnecessary confusion, back and forth communications and thus loss of time, money and man-power. Ignoring this right now can even pose a threat to our agency’s reputation. So it is highly recommended that Agency XYZ employ appropriate data curation practices and perform canonicalization of each and every data set from both the old system and the new system in order to identify and address any discrepancies or data/information loss scenarios.

While analyzing the data in both the systems it has also been noticed that metadata creation and documentation has not been given adequate priority during the data migration. Creation and management of metadata is an unavoidable step in data strategy that can influence the success rate of any data migration to a great extent. Without metadata, sharing and reusability of data can be a challenge and compatibility issues can arise during data exchanges and integration with internal and/or external systems and entities. Considering how challenging it was to examine a small chunk of migrated data in the absence of metadata, there is no question that the time ,money and other resources that we devout to creating and documenting metadata during the data curation process is going to be a great investment for the future of our organization and its data handling and data science capabilities.

It is also important that we, as an organization, document each and every step in the data migration and curation process to ensure data provenance. As you know, it is not uncommon in an agency like ours; for data formats, systems, platforms, data owners or even the data storage systems to change in no time. Sometimes changes happen to keep up with the evolving customer needs and also to take advantage of the most modern technological advances. Data provenance , if adopted, is going to make sure that every intermediate data curation and transformation process is documented in detail, so that the agency can always trace back to the initial point or data record where we started and derive insights as needed. From a data scientist’s perspective, it is equally important to have access to the properties of raw data or intermediate data as it is to have access to processed data on which machine learning or analytics can be done.

Last but not the least , I would also like to make use of this opportunity to request that Agency XYZ include data curation as a mandatory policy that needs to be part of any data migration process – big or small – including the current data migration that we have discussed in this memo. This will ensure that our agency’s data handling capabilities remain robust , benefit all our departments and above all ensure customer satisfaction thus securing the success of the organization as a whole in the years to come.

As a data scientist committed to ensuring the success of this migration, I once again strongly recommend that Agency XYZ reinstate and fund the data curation activities with immediate effect to ensure the success of the data platform migration, business continuity and to make sure that we are building a next-generation platform that supports data science and the endless data analytic capabilities that it can offer in the future. I will be glad to meet with you if a discussion is needed and can answer any questions pertaining to this. Hoping that this request will be considered favorably.

Thank You.