

MEMO

1. 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.

Total Word Count:- 538 Words 3,326 Characters

To: Director
From: Bollam Raja Shekar (bollamr2)
Date: November 30th, 2020
Subject: Why is Data Curation Services Important.

I would like to bring to your notice about the undergoing studies on data curation and its services in the context of data science. I also humbly urge you to continue funding in this area of study.

Data curation helps researchers manage, preserve, maintain and add value to data through-out its life-cycle, with the goal of providing for its re-use over time (Digital Curation Center, 2010). In current technology era, data is almost everything. Every electronics hardware or a software deals with understand the data as an input and produces output in terms of data in one way or the other. It has become very crucial to understand various forms of data.

Data representation at a conceptual level and implementation level doesn't have any concrete form of representation or guidelines to follow. Hence, data management has become one of the critical aspects of technological advancements. Therefore, having an effective and efficient data curation services has become important to pave way for cutting edge technologies in the future.

Data Curation starts with a "data set". These data sets are fundamental building blocks for data curation activities. Simplistically, the stages of data curation can be listed as Identifying, Cleaning, Transforming and Identifying. Identifying is the process of identifying different data sets or data sources that are of our interest in a given context. Cleaning is the process in which we remove not so required data and normalize the data to certain standard that we deem important for certain activity or a requirement in a given context. Transforming is the process in which we convert the normalized data into the data that is required as an input data for another process.

In this memo, I would like to emphasize the importance of Metadata and Provenance in the data curation process.

Metadata is *data that best describes the data*. It is a way of referring to the data that supports the discovery, management and understanding of other data and information. Capturing and maintaining correct metadata is considered key to the reuse and preservation of digital objects. There are a wide range of metadata schema and standards developed based on the requirements such as DDI (Data Documentation Initiative), IEEE LOM (Learning Objects Metadata) and many more.

Provenance is defined as the origins, ownership and custody of a data. Since data-sets are reformulated and reworked to create a new data, provenance is important to trace back the origins of the modified data sets. Every data set during the data curation process need to address the requirements of provenance as it guarantees the data creators accountability for their work and provides a chain of information where data can be back-traced to the creators enabling the researchers and others to adapt the data for their own purposes.

In conclusion, I strongly believe that a thoughtfully designed data curation process and Services provided by this process are one of the building blocks for technological break-through in data sciences now and more in future. I strongly urge you to continue funding this work that would allow for easier access and use of data and at the same time improving its quality.