Data Management Plan – Part III

Loic Niragire

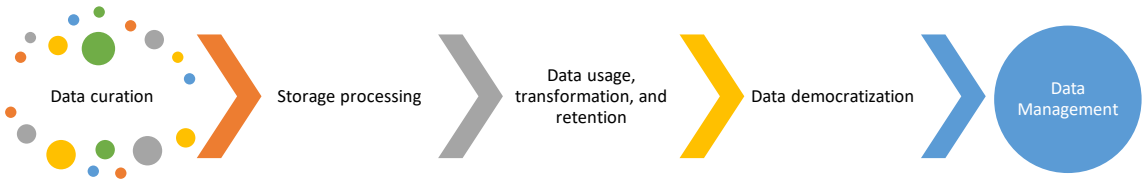
TIM-8130 v4: Data Curation

Dr. Ulrich Vouama

April 3, 2022

I propose a data democratization plan for my research team in part three of my data management plan. Part two of this plan dealt with data usage, transformation, and retention provisions. Meanwhile, I explored the role of architecture in the storage process, the purpose of data modeling, and data cataloging in part one. Figure one highlights these milestones leading to a complete data management plan starting from the data curation process. The primary goal of this data management plan is to support a team of researchers conducting studies on environmental factors and their effects on pulmonary diseases.

Figure 1 Data management plan milestones



I describe best practices for a data democratization plan that facilitates data access to researchers. Moreover, I discuss data culture transformations required for a successful data democratization initiative. Finally, I discuss policies and procedures to ensure that everyone has access to needed data, can understand it, and correctly use it. Data democratization often suffers from today's outdated models, which I will discuss later, namely the data silo and the IT ownership model. Whereby data access is constrained within business units or requires special requests through the IT department, thereby limiting data flow throughout an organization.

Best practices for data democratization

Seven steps constitute a well-thought-through democratization plan, where each step aims at answering a specific question, as illustrated in figure two below. Properly addressing these questions requires an enterprise-wide metadata view that connects organizational units, processes, technology, and data (Rose, 2012). The metadata helps people understand the data and therefore expedites the decision-making process. Thus, data democratization is more than granting data access. Equally important are the tools to facilitate its usage across the organization.

Figure 2 Data democratization steps

Data culture transformations

Earlier, I briefly discussed two standard outdated democratization models in practice today. The IT department acts as a data gatekeeper in the first model, where access is tightly controlled through approval steps. The second model is comprised of data silos within the organization, therefore lacking a company-wide view of available data and its usage. I would like to add one more cultural challenge to this list, born out of security concerns. It is the culture of a "need-to-know" basis, where data is tightly controlled, sometimes in silos, under the premise that only those who need to know should have access.

The thought of freely granting data access to all users across an organization is a security threat to be avoided at all costs. Neither is this what data democratization advocates, but rather to have the correct data available to the right people at the right time. Moreover, this data should be easily understood and in the most appropriate format for the task. To achieve this model, the first cultural change required is the understanding that all data belongs to the organization rather than any department.

To address security concerns imposed by data access, the cultural transformation required will revolve around the value we place on our metadata collection, maintenance, and sharing. The idea is that metadata should be highly regarded and managed just at the data it describes (Tang, Wu, & Li, 2019). Therefore, allowing the organization to map its data interdependencies and relations across the organization without sacrificing security. This transformation may prove daunting as most data generated from internal processes are often considered a by-product and dismissed altogether. Collecting and managing this metadata

requires a different perspective on the processes involved and data relationships across the organization.

Policies and procedures

Access to data alone does not constitute a well-designed data democratization plan. There is a need to interpret shared data and use it correctly. This section discusses policies and procedures that will ensure such attributes. I first discuss policies controlling data access and permission to keep security concerns central to our democratization plan. Since all data access is managed through well-defined Web APIs, as presented in the initial data management plan, the first policy is that all requests are granted through a centralized identity management system.

The second policy demands the implementation of an authentication and authorization mechanism on all services to control who has access and grant permissions. One benefit of this control mechanism is the granular control it affords the organization, such as the ability to grant or revoke any desired access at any given time. It also acts as a central view for all data-related access and permissions in the entire organization.

To improve ease of usage and data interpretation, all processes generating or modifying collected metadata will be required to undergo periodic review to ensure consistency with the company goals and data mission. This will guarantee that our metadata remains useful and can be accessed as needed. More importantly, this process will highlight the kind of metadata we need to be collecting while continuously exercising its effectiveness (Doctor & Joshi, 2021). As business needs change over time, our metadata needs likely

change accordingly, which requires periodic assessment to align our practices against the business needs.

Lastly, we want to increase data visibility throughout the organization to facilitate collaboration and foster a greater sense of data ownership. Therefore, we will implement a policy requiring all dataset publications to include full access to associated metadata. More importantly, referenced metadata should highlight data relationships with the existing datasets.

References

Doctor, G., & Joshi, P. (2021). Empowering Cities through Open Data-Open Government Data Initiatives in India. *International Conference on Theory and Practice of Electronic Governance* (pp. 352-361). ACM.

Rose, D. E. (2012). CloudSearch and democratization of information retrieval. *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 1022-1023). ACM.

Tang, N., Wu, E., & Li, G. (2019). Towards Democratizing Relational Data Visualization. *Proceedings of the 2019 International Conference on Management of Data* (pp. 2025-2030). ACM.