Hi Mr. A,

Welcoming again! I am so pleased to have you in the team and very excited for us working together. I understand joining new organization might be at first overwhelming. Therefore, I’m here hoping to be a piece of resources to bring you up to the momentum of our work.

Upon writing this for you, I got assigned to a project to access the quality of our data transfer and ensure the equivalence of our data when integrating to new system. During the process, I get to understand our data better and discovered some high level of inconsistency limiting our ability for a smooth data’s transition. As our organization continue to grow, large amount of data remains flowing in requiring a lot of transformations. Crisis in identifying heterogeneity along with data independent problem can happen anytime if we don’t have the way to fully understand, improving and maintaining our data. Therefore, a throughout solution is needed and as for now Data Curation is what we are looking at with the focus on metadata serve, preservation system and accurate workflow.

“Data Curation? Meta Data? Provenance? Etc”. You might wonder what they are and how they could benefit our work right now…

Data Curation can be explained from the concept of Data Science. It is belief to be the combination of two items: Data Analytic and Data Curation. While Data Analytic means getting knowledge of data using specific technique and queries, data Curation is ensuring data is found and worked efficiently and reliably. Some important factors of Data Curation are documentation, validation, transformation and integration.

When the amount of data is continuingly increasing, it will later need to be evolved into metadata. Metadata in short is “data of data” that can provide more granular version of data. *One example of meta data could be the information of a complain that we have. Each complain will includes some more information: who complains, when did the complain come in, what is our response, etc*. By designing and maintaining well-structure metadata we could achieve documentation factor of Data Curation. It tells you where data come from, where it locates at in which system enhancing user’s understanding. It also helps interpreting data with familiar attribution’s semantics, allows data quality and integrity, avoids illegal use and support smoothing the integration process.

In addition to the need of metadata, a preservation system is necessary to be in place ensuring reliability communication for our future use. Continuing make a lot of copies for our data, keeping them updated with new formats (if there is), maintaining software and convert new data set to standard format of preservation.

We also need to have a workflow system either a manual system that we can develop or a specialized workflow system in support the nature of data transformation. That to help us to keep tracks of what is going on with each transition, identify what is being done or not being done, what are inputs, what are output.

I strongly believe data is one of our biggest assets. Piling up tons of our asset without a system in place is not practicing the fundamental Return on Asset principle. We then need a way to value our data, finding the potential and keep on good practice with maintaining and transforming it. Data Curation, which is now a vocabulary in Big Data management, could be a practical solution that we have. Therefore, I hope while transitioning to this new role, you will keep this idea in mind and continue allowing us to keep a good hand on our data asset.