Joseph Annand

DSE5004

Position Paper #1

### Organizational Challenges

The key organizational challenges in translating data into information from which knowledge can be extracted are centered around data availability through access and sharing. A data scientist must question during any project about what data pipelines and data governance policies are in place so that they have sufficient resources to complete the project in mind. In order to translate data into information that will become meaningful and worthwhile to business decisions, high-quality data must be readily available and easily attainable. One issue that may arise in data availability is the storage of cross-functional data. For example, data from different departments, like finance and sales, may be stored in separate locations, making a difficult for analysts to locate and retrieve data necessary for a particular assignment. Additionally, how this data is catalogued is another factor in data availability. If, for example, the company does not have standardized file naming conventions, it may be difficult to locate specific files or data sets (7 Best Practices for Successful Data Management).

Data security is essential for any company, as data breaches are a significant financial risk; however, poor governance policies that inhibit access to data may have a negative impact on analytics and, thus, business efficiency. If data security protocols are too strict, data scientists and analysts will face challenges in retrieving data, which will waste time and money (What is Data Management?) Companies need to strike the right balance between convenience and security by allowing appropriate individuals access to data while maintaining security through de-identification efforts and protective data storage and transfer methods, such as encryption.

### Data Quality

Data quality presents a major issue in the data-knowledge continuum. Poor-quality data makes it more difficult for analysts to develop information that will provide valuable insights for business decisions. To begin, an understanding between analysts and those creating data entries needs to be established. Whether it is external data that researchers found and added to the database or it is other company employees entering internal data, data scientists and analysts need to communicate with whomever it is appropriate the needs and expectations they have in data collection and storage so that they can save time and more effectively complete analytical tasks (Redman, 2020). If not, the analysts may find themselves correcting mistakes, like formatting errors, or needing to request additional information from the individual who originally entered the data because crucial features are missing.

Additionally, there is a wide variety data quality features that must be accounted for, such as accuracy, completeness, timeliness, consistency, and data granularity and relevance (Jones, 2023). Without taking these features into account, it will be harder for data scientists to translate data into information, and it puts the company at risk of making poorly informed decisions. Data validation rules and methods should be in place to ensure accuracy; data sets should be complete by handling missing values and merging with other relevant data sets; and data must be kept up to date so that business decisions are informed by relevant information. When data is too complex, or has excessive granularity, the transfer of information to knowledge may be placed in jeopardy: data that contains unnecessary detail will make it difficult to draw meaningful insights that translate to business actions (Jones, 2023). The same is true if the data is not detailed enough, so finding a balance is crucial to the data-knowledge continuum.