Roundtable on “Data Management Issues in Digitization of Supply Chains”

*** Executive Summary ***

“Digitization of supply chains” is a buzzing phrase these days as companies are trying to figure out the implications of the emerging technologies such as Big Data, Machine Learning, Internet-of-things, Blockchain, etc. for their supply chains. Behind all this hoopla lies a rather mundane but important issue: management of data!

If companies are to prepare for the upcoming technological revolution (often termed as the “Fourth Industrial Revolution” or “Industry 4.0”), they need to answer some fundamental questions about data – what data to collect, what sources to collect it from, how frequently to collect it, how to store it, how to make right data available to the right user, how to help potential users understand the data correctly, how to check and ensure validity of the data, and so on.

Malaysia Institute for Supply Chain Innovation (MISI) hosted a roundtable – with its partners at the MIT Center for Transportation & Logistics (CTL) – to explore such “Data Management Issues in Digitization of Supply Chains”. Thirteen IT and supply chain professionals discussed these issues in the one-day roundtable on November 13, 2019. This report presents key findings of this roundtable.

Participants

Participation in the roundtable was by invitation only. All companies invited to the roundtable were handpicked by MISI and MIT CTL because of the diversity of their expertise and experience in the topic, and the insights their participants were expected to bring to the fore. Thirteen individuals from nine organizations participated in the roundtable discussion. They came from a diverse collection of sectors of the economy, such as aerospace, chemicals, computing hardware, computing software, e-commerce, electronics manufacturing services, oil & gas, original design electronics manufacturing, and research-focused organizations.

Roundtable sessions

Discussion at the roundtable was organized into four sessions (below). Each session began with a presentation by one of the roundtable participants and was followed by a discussion among all participants, facilitated by Mr. Jim Rice (MIT CTL) and Dr. Shardul Phadnis (MISI).

- Session I: Managing Transformation and Change through Use-Case-Based Prioritization
- Session II: Future State Data Governance Structures & Policies
- Session III: Collecting the Right Data
- Session IV: Data Preparation & Formatting

The discussions were rich and produced interesting insights. Two note-takers—both supply chain professionals, each with close to a decade of industry experience, and current students in MISI’s
Master of Science (Supply Chain Management) program—took detailed notes of the discussions, according to the guidelines provided by Dr. Phadnis. The facilitators, Mr. Rice and Dr. Phadnis, themselves noted the key points during the discussion. The analysis of this data yielded ten key lessons for the roundtable across the four sessions.

**Key lessons**

**Lesson 1: Choose appropriate data storage strategy for the organization**

Data management starts with an overarching strategy for data storage in a multi-unit organization. An organization may use localized data strategy that allows individual units most control over data collection and sharing, a centralized data strategy that allows different business units to work with a single source of truth, or an intermediate federated model that centralizes control and management of the data infrastructure but allows business units to control data collection and even own the data lake(s).

**Lesson 2: Practicing sustainable data management requires careful organization design**

A sustainable digitization initiative requires a careful organization design. Although different organizational forms are possible, a critical requirement for success is that requests for data projects are initiated according to business needs (by business units) and implemented while adhering to good data management and governance practices (which may be overseen by a dedicated IT team). As in any significant initiative, the digitization initiative also requires a strong support and continued involvement of the organization’s top management.

**Lesson 3: Potential pitfalls include underestimating data education and availability of right talent**

A set of key pitfalls relate to the human element of data management. An obvious pitfall is to expect data management to solve all problems, including those resulting poor organizational processes. Companies may also underestimate the amount of effort required in educating business unit associates on data definitions to ensure that they understand how the data they collect gets used for making decisions and what data is used in the decision-making aids they use. Finally, members of a good data management team need to possess the twin qualities of understanding the business need while being skilled in the data science and management techniques.

**Lesson 4: A multi-tiered data governance framework is necessary to manage data quality**

A four-tiered framework, in which the lower tiers provide the foundation for higher tiers, is helpful for ensuring data governance. The lowest tier pertains to master data management to ensure that the master data files have correct data, are updated periodically and systematically, and are the single source of relevant information. The second tier involves the use of data catalog and glossary to provide
definitions of and relationships among different data in the organization’s data lake. The third tier is a data quality monitoring platform that can continuously monitor the data entering the organization’s data lake and highlight potential instances of poor data quality for further action. The last tier—currently, aspirational—uses machine learning to monitor and correct data automatically and advise data scientists about the potential issues that may not be fixed automatically.

**Lesson 5: Data normalization is essential, regardless of the choice of data architecture**

Data normalization is the practice of mapping fields with similar data to one field, where the similar data will be stored, to avoid redundancy. Subsequently, this field is given an appropriate name so that it can be understood by different parties in the organization. A data dictionary is used to list definitions and descriptions of the normalized fields in the organization’s data lake. Data normalization is an essential task regardless of the data architecture used by the organization, and requires collaboration between business units and the MIS/IT team.

**Lesson 6: Know the metrics of data quality, and what poor-quality data looks like**

Data quality is paramount for a digitization initiative to succeed. Poor quality data has at least four attributes: different terms are used to refer to the same information (misleads users to believe different data pertains to different information), duplicate records (creates confusion about which data to trust, increases cost of data storage), different interpretations of the same data (leads to incorrect interpretations and poor decisions), and non-standard data (makes interpretation difficult). Such problems can be mitigated by using key performance indicators (KPIs) to measure data quality, such as accuracy, completeness, conformity, and consistency.

**Lesson 7: Companies need to decide which data to collect, process and store**

Data collection begins with data entries, either automatic or manual, into structured databases or by harvesting unstructured data from predefined sources such as social media, video feeds, or weblogs. The collected raw data needs to be cleaned, classified, and identified before it is made available for analysis. Business units’ requests for collection of new data need to be scrutinized to ensure that the new data is necessary and complies with good data governance. In addition, companies also need to decide which data to store. As the volume and velocity of data collected increases, the answer to whether and which data to store requires resolving tradeoffs between costs of data storage and benefits of granularity of the data available for fast analysis.
Lesson 8: Unstructured data has its own set of challenges and benefits

Much of the data gathered by organizations today in their IT systems is ‘structured’. A majority of the companies undertaking supply chain digitization initiatives are focused primarily on exploiting structured data. However, unstructured data (such as that scraped from social media, video feeds, blogs, etc.) can provide novel insights. However, unstructured data comes with its own challenges. Such data may need sophisticated processing before being made available for analysis (such as natural language processing to parse multi-language comments, emoji’s, audio/video feeds, etc.). Organizations also need to decide how much of the unstructured data to store: storage of data such as video feeds can be expensive, and this expense needs to be traded off against the benefits of archiving such data.

Lesson 9: Longitudinal, end-to-end data may be necessary for a particular analysis

Data processing is the last key step in data management before the data is made available for analysis. The objective of this step is to store data to facilitate accurate analysis. Data analytics to gain insights may require using data collected longitudinally as well as for the end-to-end process. Therefore, companies may need to process and store time-series data collected at different steps in a product’s end-to-end supply chain. Such data is processed to identify and correct duplicate entries, missing records, and erroneous readings.

Lesson 10: Data management needs various skills; talent is hard to retain with money alone

Data management requires a variety of skills that are unlikely to be found in one person (e.g., statistical modeling for data scientists, data processing for data engineers, business acumen business intelligence analysts). Most companies experience difficulty recruiting talented employees to fill various data-related roles. Once found, retaining good employees is also a challenge. Although there are no definitive solutions, the roundtable participants concurred that money alone is not adequate to retain good employees. Talented employees need challenging problems to work on, and have appropriate tools, data, and the freedom to tackle them. Companies may need to ask if they are IT-driven companies, and not just the traditional industry-focused companies, if they are to recruit and retain talented employees to fulfill various data management roles.

This report was prepared by Dr. Shardul Phadnis at Malaysia Institute for Supply Chain Innovation. For further information, contact sphadnis@misi.edu.my