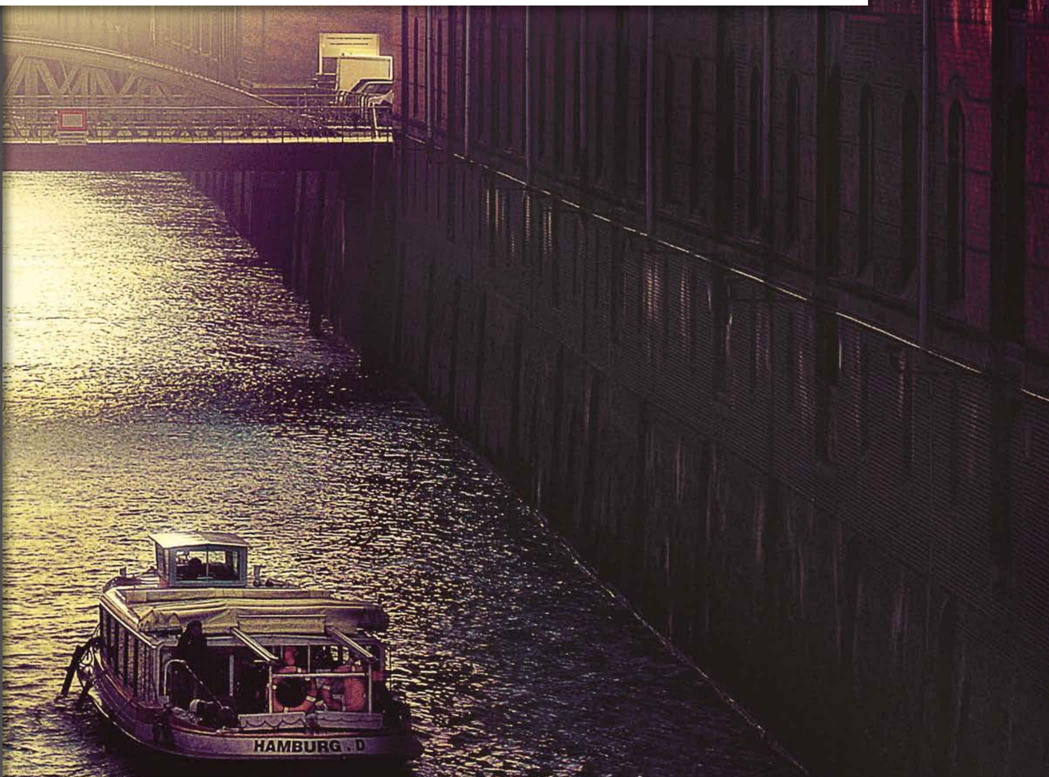
Proceedings of the Hamburg International Conference of Logistics (HICL) – 23

Johannes Zrenner, Ahmad Pajam Hassan, Boris Otto, Jorge Marx Gómez

Data Source Taxonomy for Supply Network Structure Visibility



Published in: Digitalization in Supply Chain Management and Logistics
Wolfgang Kersten, Thorsten Blecker and Christian M. Ringle (Eds.)
ISBN 9783745043280, Oktober 2017, epubli

Data Source Taxonomy for Supply Network Structure Visibility

Johannes Zrenner[1], Ahmad Pajam Hassan[2], Boris Otto[1], Jorge Marx Gómez[2]

1 – Technical University of Dortmund

2 – Carl von Ossietzky University of Oldenburg

The supply network structure of manufacturers is complex and non-transparent. In order to achieve a higher visibility and consequently increase the performance, the existing lack of data has to be closed. This paper answers the questions, how to identify, describe and compare suitable data sources for an end-to-end visibility. Following the design science research process, two artifacts are developed based on conceptual-to-empirical approaches. The initial conceptualizations result from literature reviews. The conceptual representation of supply network structure data sources clarifies the relevant data entities and attributes. It supports the identification process of relevant data sources. The data source taxonomy (i.e. classification scheme) describes data sources using fourteen dimensions and up to four potential characteristics. It assists a standardized description. Both artifacts are demonstrated in case studies with German automotive Original Equipment Manufacturers. The findings add to the knowledge base of supply network visibility with a focus on the network structure. A large part of the existing literature about supply chain visibility is too vague on the data perspective. Therefore, this paper closes an important gap regarding the supply chain digitalization by introducing two applicable results, which enable a new course of action for practitioners and researchers.

Keywords: Data Source Taxonomy; Supply Chain Visibility; Supply Network Structure; Design Science Research

1 Introduction

In the age of globalization, cost pressure and the customizing requirement, companies are constrained to be focused on their main capabilities. Therefore, make or buy decisions are followed by outsourcing business processes, which makes supply networks more complex and dynamic (Tang and Tomlin, 2008). For handling the dynamic of supply networks and reduce the uncertainty, it is necessary to make supply network relations visible (Christopher and Lee, 2004). Thus, visibility becomes a key topic in Supply Chain Management research (Yu, Yan and Edwin Cheng, 2001). With more visibility in supply chains, the performance improves (Pidun and Felden, 2012) and decision making processes get supported (Kulp, Lee and Ofek, 2004).

From the perspective of a manufacturer, visibility of all supply network members is only given for the first tier supplier (Basole and Bellamy, 2014), because of direct business relationships. Data about them is internally stored in databases. After the first tier supplier, the upstream supply chain gets increasingly unknown (Christopher and Lee, 2004). Missing interconnections between data silos among business areas, supply chain members and other external data sources are the reason for the lack of end-to-end visibility. This issue can be solved by linking disparate data sources. The various available data sources are highly heterogeneous, own different characteristics and contents (Rozados and Tjahjono, 2014). For that reasons, it is difficult to identify and select the suitable data sources for linking processes. Furthermore, a general understanding of the data sources is missing, which is needed to negotiate about data source sharing with supply chain members. For example, potential data sources are owned by a first tier supplier, logistics contractors or data providers.

The goal of this paper is to enhance the knowledge about data sources for supply network structures and thus support linking processes for practitioners and researchers. According to that, the first research question (RQ) addresses a general data model, which helps to identify data sources with a suitable content. The data model clarifies the entities and attributes, a data source about the supply network structure has. The first research question is:

RQ1: How does a conceptual representation for supply network structure data sources look like?

Following the identification of available data sources, a comparison has to be conducted. The goal is to find out, which are the most suitable data sources from an effort-benefit point of view, for bridging data lacks within the supply network. For that reason, a taxonomy (i.e. classification scheme) is needed, which standardizes the description of the data sources and makes them comparable. Accordingly, the second research question is:

RQ2: How does a taxonomy of data sources for supply network structures look like?

This investigation contributes to the research of information systems. It follows the design science research (DSR) process from Peffer et al. (2007). It is an iterative research and design process. The process combines scientific and practitioner's knowledge for designing the artifact.

The remainder of the paper is structured as follows. Section 2 lays out the related work of concepts for supply chains and data source taxonomies. Section 3 presents a conceptual representation for supply network structure data sources. In section 4 the data source taxonomy is established. Section 5 describes the application of the conceptual representation and the taxonomy. Section 6 demonstrates the conceptual representation and the taxonomy through two case studies with German Original Equipment Manufacturers (OEMs). Section 7 summarizes the contributions to research and practice as well as showing fields for future research.

2 Related Work

2.1 Concepts for Supply Chains

Supply Chain Management is focused on managing the whole supply chain from the raw material producer to the end customer (Harland, 1996). The intentions of supply chain integration are improvements on performance and operational figures (Ramdas and Spekman, 2000; Frohlich and Westbrook, 2001; Rosenzweig, Roth and Dean, 2003; Cagliano, Caniato and Spina, 2004).

This investigation addresses the supply network from n-tier supplier to its Original Equipment Manufacturer in a different way from most other researches. They are focused on processing data from the interaction between supply network members (Levary, 2000; Zhao, Xie and Zhang, 2002) and data sharing (Bowersox, Stank and Closs, 2000). The object of this investigation is to analyze different data sources about business relations in a manner, that it can be used for increasing the visibility of supply network structures. In this context, Mukaddes et al. (2010) propose a conceptual information model for Supply Chain Management and integration. The model deals with the integration of supply chain members and focuses on the flow of information between the chain members. Grubic and Fan (2010) present a study of state-of-the-art research in supply chain ontology. They identify outstanding research gaps and six supply chain ontology models are identified from a systematic review of literature. In the context of decision support in global supply chain, Wang, Wong and Fan (2013) build an ontology for steel manufactures, which represent eleven main classes and the associated attributes.

Despite existing contributions, there remains a need to specify a dedicated ontology as a model for conceptualizing supply network structure data sources.

2.2 Data Source Taxonomies

In the generation of big data, there are a lot of data sources relevant for Supply Chain Management. Leveling, Edelbrock and Otto(2014) propose an overarching model of data categorization. It classifies data in nucleus data, community data and open big data. For this, the following characteristics are considered: fuzziness, volume and change frequency.

A review from Rozados and Tjahjono (2014) identifies 52 data sources and classifies them to a taxonomy. The main characteristics for this classification are volume, velocity and variety. The characteristic variety is divided into structured data, semi-structured data and unstructured data. For example, core transactional data like transportation costs, origin and destination are structured data with small volume and velocity. Semi or unstructured data like weather data or machine-generated data have a high volume and velocity.

Otto, Abraham and Schlosser (2014) propose a morphology, which makes the relevant characteristics of the data resource in networked industries transparent. The morphology is based on four case studies and has eleven dimensions with two to six characteristics. Even though this morphology is pretty detailed, it is specific to the data resource. Data resources entail all databases in a company.

Given this scarce knowledge base about data source taxonomies, there is a clear need to investigate this topic further.

3 Conceptual Representation of Supply Network Structure Data Sources

Since there is no common standard for the vision of an integrated supply network through data flow integration (Loh, Koh and Simpson, 2006; Patnayakuni, Rai and Seth, 2006), it is necessary to find new approaches to make the supply network structure visible. For this purpose, the first step is to define the relevant aspects of supply network structure data sources. Therefore, a conceptual representation based on domain specific literature is developed. The authors identify relevant literature, extract fundamental supply network data features and create an entity relationship model following the framework of Chen (1976). By space limitations, only the most relevant components of the model are explained in detail.

"The membership in a network and its organization vary for a given product and over time" (Choi and Hong, 2002), the attribute "valid time" is considering this fact and makes relations time sensitive. It is assigned to the relation "deliver" and "demand". In respect to the data and product flow, it is also necessary to differentiate between several kinds of functional areas within an organization and between supply network member organizations (Basole and Bellamy, 2014). Following that, the entity company owns an attribute representing the type of location. Examples for locations types are supplier, manufacturer, warehouse, distributor and retailer. Moreover, to the entity transferred product an attribute is assigned, which classifies its type, for example raw material, parts and components. Companies within a supply chain are using applications which are based on different ontologies. That leads to inconsistent terms and semantics which have a negative influence on the interoperability for supply chain integration (Ye, et al., 2008). According to the eCl@ss standard (eCl@ss, 2017), the entity product could own different designations, for example identifier, preferred name or short name. Those designations could even differ between languages. Considering the variety in semantics, both entities "product" and "company" own an "alias" feature. This feature makes it possible to express different semantics, which are used by their data source.

Figure 1 shows the designed representation. It provides the relevant aspects of supply network structure data sources in a formal way. Furthermore, it is used in the taxonomy application (section 5) to identify potential data sources.

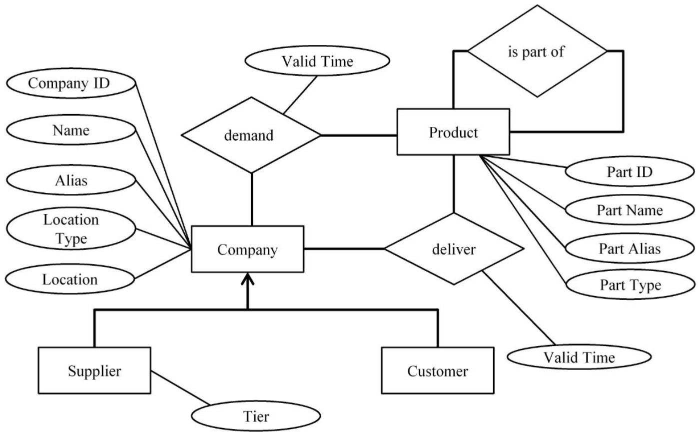


Figure 1: Conceptual representation of supply network structure data sources

4 Data Source Taxonomy

4.1 Background

A taxonomy is a classification scheme of objects from a specific research area. It supports scientists and practitioners to understand and analyze complex domains (Nickerson, Varshney and Muntermann, 2013). A morphological box structures the identified dimensions and characteristics of the investigated objects. Therefore, the morphological box is a suitable presentation of a taxonomy.

For the taxonomy development in information systems, Nickerson et al. (2013) propose a method with two different approaches. The empirical-to-conceptual approach starts with identifying a subset of objects and determining the characteristics. After that, the characteristics are grouped into dimensions and the taxonomy is created. The first step of the conceptual-to-empirical approach is the conceptualization of the characteristics and dimensions. Step two is about the examination of objects with these characteristics and dimensions. The third step is to create the taxonomy.

4.2 Taxonomy Development Process

The authors follow the method of Nickerson et al. (2013) and choose the conceptual-to-empirical approach. For researchers with significant knowledge about the domain, it is the recommended approach (Nickerson, Varshney and Muntermann, 2013). The initial conceptualization of the dimensions and characteristics is based on the researcher notions about data sources as well as a review of the literature. The authors scour scientific databases (IEEE Xplore, Web of Science, ACM DL, Emerald Insights, ScienceDirect) for the following search terms: “data taxonomy”, “big data in Supply Chain Management” and “Supply Chain Visibility”. Two demonstration iterations in a business environment (section 6) are used to examine data sources with the identified dimensions and characteristics.

The results of the described process lead to the taxonomy of data sources for supply network structures (figure 2).

Dimension	Characteristics			
Data Source Availability	Internal	External – Closed		External – Open
Data Source Interface	Internal Interconnection	Traditional EDI	Web Services	Offline Data Dump
Data Source Pricing Model	Volume-Driven	Time-Driven	Unique	No
Data Aggregation	Resource	Database	Record	Item
Data Occurrence/Update	Stream	Event-Driven Batch		Time-Driven Batch
Data Ownership	One Legal Entity	Community		Public
Data Structure	Structured	Semi-Structured		Unstructured
Data Format	Proprietary		Open	
Intra Data Standardization	Value	Semantic	Syntax	No
Inter Data Standardization	Value	Semantic	Syntax	No
Data Currency	Forecast	Up-To-Date		Outdated
Data Completeness	High	Medium		Low
Data Accuracy	High	Medium		Low
Data Sharing	Proprietary	Free		Open

Figure 2: Taxonomy of Data Sources

4.3 Taxonomy Details

Each of the fourteen dimensions has two to four potential characteristics, which are not exclusive but listed in the same row. That means, one dimension can have one or more applying characteristics.

Data Source Availability: If a data source is internal or external, is the main differentiation between data sources. From an enterprise perspective, an internal data source is available in the company's IT infrastructure for example, a database of the ERP system. External data sources are not available in the company's IT infrastructure and can be divided into closed and open data. The difference is the charged or free access to the data source. When the data has to be either purchased or licensed, the data source is called closed (Leveling, Edelbrock and Otto, 2014). Open Data is freely available on the public internet and can be used without technical, financial or legal barriers (Murray-Rust, et al., 2010).

Data Source Interface: For end-to-end visibility of supply network structures, data from different sources has to be accessible from a single point of access. Therefore, an interconnection of data sources is required. Internally available data sources can be distributed across business areas (Rozados and Tjahjono, 2014) and thus an internal interconnection is sufficient. Between the OEM and the supplier normally exists a traditional Electronic Data Interchange (EDI) connection for transactional data (Zilbert, 2000), which can be the connection for those kinds of external sources. Web Services are a cross-platform way to connect external data sources (Li, Sun and Tian, 2015) from a supplier, data provider, or Open Data. A Data dump is a snapshot of data and is often used when IT infrastructure is limited for a single or prototypical data transfer, like for example an Excel file sent by email.

Data Source Pricing Model: For each of the four used characteristics, one example of a potential supply network structure data source is named. The pricing model from Achilles (2017) is based on the number of used product codes. Therefore, Achilles is an example for a volume-driven pricing model. Panjivia (2017) has a time-driven model with a monthly and annually option. IHS (2014) offers reports for a unique payment. When linking data sources within a division of a company, normally no pricing model is used.

Data Aggregation: Data aggregation describes the aggregation level of the data source. An instantiation of an attribute of a data object is called data item. A data record is the instantiation of a data object. Data records constitute database tables.

The highest level of aggregation is data resource. The data resource includes all databases of a company (Otto, Abraham and Schlosser, 2014).

Data Occurrence/Updates: Data occurrence is distinguished in batches and streams. Examples for streams are the social media streams or the shop-floor data streams (Otto, Abraham and Schlosser, 2014). Batches are sets of data records and no continuous data flow like streams. Therefore, data updates are usually for batches. There can be an event or time-driven trigger for updates.

Data Ownership: If data is a private good, it is owned by a legal entity like a company. Data as a club good is owned by a community and public data goods like addresses are available for the public (Otto, Abraham and Schlosser, 2014).

Data Structure: Data without an identifiable structure or a data model is called unstructured data. For example, photos or narrative text are unstructured data (Kubler, et al., 2015). Semi-structured data like XML or machine-generated log files have an irregular and flexible structure, but cannot be processed in relational databases efficiently. However, ERP transaction data or origination and destination are structured data and can be straightforwardly transferred to a relational database (Rozados and Tjahjono, 2014).

Data Format: Open data formats, such as HTML, XML, JSON, RDF or CSV can be used by anyone, unlike proprietary formats. Binary data from machine to machine communication is one example for proprietary formats (Kubler, et al., 2015). PDF or DOC are other common proprietary data formats.

Data Standardization: Data can be standardized on the syntactic level, on the semantic level and on value level. Standardization on a value level standardizes the possible values of a data item (Otto, Abraham and Schlosser, 2014). Semantic heterogeneity through homonyms or synonyms leads to different meanings or interpretation problems. A homonym denotes different terms and a synonym denotes the same term as another notation (Bergamaschi, et al., 2011). The demonstration of the taxonomy (section 6) leads to additional characteristics. If the data source is not standardized at any level, the characteristics “No” is necessary. Moreover, the demonstration results lead to a differentiation between “Intra Data Standardization” and “Inter Data Standardization”. Intra refers to the standardization inside of a data source. On which level the standardization of the data fits from one data source to other ones, is covered by “Inter Data Standardization”.

Data Quality and Data Value: Data quality is essential for data analytics because the accuracy of the analytics methods depends on the data which they are based on. That means the results can only be as good as the input data. The intrinsic data quality dimensions are (Hazen, et al., 2014):

— Accuracy: Does the data equivalent to their corresponding real values?

— Timeliness: Currency (length of time since the record's last update) and Volatility (frequency of updates)

— Consistency: Matches the data regarding format and structure?

— Completeness: Is the data complete or is there data missing?

Extended literature considers the value of information and neglects the value of data. Data is used to generate information and therefore scholars do not distinguish between information resources and data resources. Data quality affects data value and the study of Ahituv (1989) identifies influencing factors for the value of information (Otto, 2015). The influencing factors for the value of information are classified into four categories (timeliness, contents, format, cost) (Ahituv, 1989). Except "Currency", "Completeness" and "Accuracy", all named data quality and data value attributes are already covered in the taxonomy. The currency of data can be distinguished in forecast, up-to-date and outdated. Completeness and accuracy are very situational attributes. Therefore, a distinction between low, medium and high is appropriate. For assessing the completeness, the conceptual representation of supply network structure data sources (figure 1) can be helpful.

Data Sharing: The shareability of data can be distinguished in proprietary, free and open. Free and open data are allowed to being shared. The difference is that the source of free data always has to be disclosed. However, it is not allowed to share proprietary data (Otto, Abraham and Schlosser, 2014).

5 Application

Prerequisite for the application of the taxonomy is the identification of data sources, which should be described or compared for linking processes. The conceptual representation (figure 1) is used for identifying potential data sources, by demonstrating the relevant data entities and attributes.

Through applying the taxonomy on the identified data source, the taxonomy is instantiated. Therefore, each dimension of the taxonomy is analyzed and the corresponding characteristics of the data source are identified. As a result of the application process, the identified characteristics are highlighted in the taxonomy or written in a new table like in section 6. An important factor for an effective application process are the involved roles. Very suitable roles from a knowledge and capability perspective are data stewards, data owners, data architects and data scientists (Otto, Abraham and Schlosser, 2014).

6 Demonstration

According to Hevner, March and Park (2004), case studies are a suitable demonstration method for artifacts. In order to demonstrate the applicability of the developed conceptual representation of supply network structure data sources and the taxonomy, the authors make case studies with two German OEMs. The data source availability is the superior and most selective data source dimension in the taxonomy. Therefore, each case study covers one data source scenario for each characteristic of that dimension (internal, external-closed, external-open).

The authors perform one workshop with each OEM. The workshops are divided into several steps. The first step is an introduction to the paper‘s topic and its goals. After that, the workshop participants discuss the dimensions and characteristics of the taxonomy. The results of that discussion are already considered in the taxonomy (figure 2). Step three is the collective decision of the data sources for the three data source scenarios. The selected data sources should have the potential for reconstructing the supply network structure of the OEM, which is determined by using the conceptual representation of supply network structure data sources (figure 1). Step four is the application of the taxonomy for each scenario. The last step is a brainstorming session in order to obtain feedback to the data model and the taxonomy. The participants and results of both workshops are described below.

6.1 Case 1 - OEM 1

The participants of the workshop are the manager of the department “Information Processes Logistics”, one data architect and one data scientist. The results of the discussion about the dimensions and characteristics of the taxonomy (step

Table 1: Scenario I - SAP MM from the OEM

Dimension	Characteristics
Data Source Availability	Internal
Data Source Interface	Internal Interconnection
Data Source Pricing Model	No
Data Aggregation	Database, Records, Items
Data Occurence/Update	Event-Driven Batch, Time-Driven Batch
Data Ownership	One Legal Entity
Data Structure	Structured, Semi-Structured
Data Format	Proprietary
Intra Data Standardization	Semantics, Syntax, Values
Inter Data Standardization	Semantics
Data Currency	Forecast, Up-To-Date, Outdated
Data Completeness	High
Data Accuracy	High
Data Sharing	Proprietary

two) are two changes in the taxonomy. The name of the dimension "Data Organization" changes to "Data Aggregation". Data Organization does not describe the corresponding characteristics well. Organization is a very general term and leads to a lack of clarity. The OEM has future supplier relations in his ERP system. Thus, the characteristic "Forecast" becomes part of the dimension "Data Currency".

In step three, the participants decide about the IT systems or rather data sources for the three scenarios, by using the conceptual representation. The first scenario is about the OEM's internal SAP MM (Materials Management) system. The Material Management module is part of the SAP ERP (Enterprise Resource Planning). It contains material, supplier and other master data. Furthermore, it contains transaction data about material flows from the supplier to the OEM. The scenario about the internal SAP MM data source is summarized in table 1.

Scenario II is about the SAP MM system of a 1-tier supplier, which is an external and closed data source. In contrast to the OEM's SAP MM system, there is master data and transactional data about 2-tier supplier, which are part of the supply network from the OEM. Table 2 includes the characteristics of the data source "SAP MM from a 1-tier supplier".

Table 2: Scenario II - SAP MM from a 1-tier supplier

Dimension	Characteristics
Data Source Availability	External-Closed
Data Source Interface	Traditional EDI, Web Services, Offline Data Dump
Data Source Pricing Model	No
Data Aggregation	Records, Items
Data Occurence/Update	Event-Driven Batch, Time-Driven Batch
Data Ownership	One Lega Entity
Data Structure	Structured, Semi-Structured
Data Format	Proprietary
Intra Data Standardization	Semantics, Syntax, Values
Inter Data Standardization	Semantics
Data Currency	Forecast, Up-To-Date, Outdated
Data Completeness	High
Data Accuracy	High
Data Sharing	Proprietary

For scenario III the open data source “Eurostat” (2017) is used, which provides databases like statistics about international sourcing activities, subsidiaries and air freight transports (table 3).

In the feedback session, the participants report a good level of satisfaction with the workshop results. In their opinion, the results provide clear insights into the different characteristics of the available data sources. The conceptual representation and the taxonomy are a useful support to identify the data sources with a high potential for increasing the supply network structure visibility. The determination of the data source characteristics presupposes knowledge and experience in the field of data science. That is the critique of the participants because data scientists are rare in this company. In conclusion, the department “Information Processes Logistics” will use the taxonomy to compare data sources for future linking projects. They have a number of project requests from operational logistics departments, which need more data about the supply network for a proactive risk management.

Table 3: Scenario III - Eurostat

Dimension	Characteristics
Data Source Availability	External-Open
Data Source Interface	Web Services, Offline Data Dump
Data Source Pricing Model	No
Data Aggregation	Records
Data Occurence/Update	Event-Driven Batch, Time-Driven Batch
Data Ownership	Public
Data Structure	Structured
Data Format	Open
Intra Data Standardization	Semantics, Values
Inter Data Standardization	No
Data Currency	Outdated
Data Completeness	Low
Data Accuracy	Medium
Data Sharing	Open

6.2 Case 2 - OEM 2

The participants of the workshop are the head of the department "Information management of Procurement" and two data scientists.

In step two the participants discuss and review the taxonomy. The outcomes are two changes. The first one is adding the characteristic "No" to the dimension "Data Standardization". This value is rarely needed in a business application focus, but it becomes important when external data sources are investigated. The second change is the partition of the dimension "Data Standardization" into "Intra Data Standardization" and "Inter Data Standardization". The intra standardization refers to standardization inside a data source. The inter standardization describes the degree of standardization over various data sources.

In step three, the participants choose three scenarios to demonstrate the taxonomy. Scenario IV is about the OEM's internal and customized visual basic application, which is based on a spreadsheet for the purpose of capacity management. It contains material, supplier and other master data. Furthermore, it contains transaction data about material flows between sub-supplier, supplier

Table 4: Scenario IV - Spreadsheet for Capacity Management

Dimension	Characteristics
Data Source Availability	Internal
Data Source Interface	Offline Data Dump
Data Source Pricing Model	No
Data Aggregation	Records
Data Occurence/Update	Event-Driven Batch
Data Ownership	One Legal Entity
Data Structure	Structured
Data Format	Proprietary
Intra Data Standardization	Semantics, Syntax, Values
Inter Data Standardization	No
Data Currency	Up-To-Date
Data Completeness	High
Data Accuracy	High
Data Sharing	Proprietary

and the OEM. Table 4 summarizes the characteristic of the Spreadsheet for Capacity Management.

The fifth scenario (table 5) is about an external data source. The market research firm "IHS" provides a collection of automotive supplier relationships called "Who Supplies Whom". It contains data about 700 supplier and their automotive components, modules and systems. The online accessible demo data source does not provide a full access (IHS, 2017). The whole data set is just available for registered customers. This is the reason why the data source is characterized as external and closed.

The sixth demonstration scenario is about an external and open data source. Online news and press releases are available on the internet and provide data about relations between companies. An example for supply network structure data in online news is the product press release on a supplier website (AGC, 2017). Table 6 summarizes the scenario about online news and press releases.

The brainstorming session in the last step of the workshop brings up the following feedback from the participants. The taxonomy clarifies the differences and similarities of public data sources and data sources, which are liable to costs.

Table 5: Scenario V - IHS Who Supplies Whom

Dimension	Characteristics
Data Source Availability	External-Closed (only demo is open)
Data Source Interface	Web Services, Offline Data Dump
Data Source Pricing Model	Volume-Driven, Time-Driven
Data Aggregation	Database
Data Occurence/Update	Time-Driven Batch
Data Ownership	One Legal Entity
Data Structure	Structured
Data Format	Proprietary
Intra Data Standardization	Semantics, Syntax, Values
Inter Data Standardization	Semantics
Data Currency	Up-To-Date, Forecast
Data Completeness	High
Data Accuracy	High
Data Sharing	Proprietary

Table 6: Scenario VI - Online news and press releases

Dimension	Characteristics
Data Source Availability	External-Open
Data Source Interface	Web Services
Data Source Pricing Model	No
Data Aggregation	Item
Data Occurence/Update	Event-Driven Batch
Data Ownership	One Legal Entity, Community, Public
Data Structure	Unstructured
Data Format	Open
Intra Data Standardization	No
Inter Data Standardization	No
Data Currency	Up-To-Date
Data Completeness	Medium
Data Accuracy	Medium
Data Sharing	Proprietary, Free, Open

Therefore, it is simpler to identify weak spots of a data source and make a decision based on objective criteria. The conceptual representation helps to focus on the essential data source content and supports the selection processes of the large number of available data sources. A way to quantify the suitability of data sources for increasing the supply network structure visibility, would be a field for future research. The quantification helps to identify the most suitable data source for a linking project with current data sources. The department "Information management of Procurement" expects a strategic advantage for sourcing and supplying processes by increasing the supply network structure visibility.

7 Conclusion

Following the design science research process, this research addresses two questions to implement a higher supply network structure visibility through the linking of different data sources. The first question, how a conceptual representation for automotive supply network structure data sources looks like, is answered with an entity relationship model. It defines the essential content, a data source about automotive supply network structure has to have. More particularly, the model clarifies the necessary entities and attributes of a data source for improving the supply network structure visibility. The second question deals with a taxonomy of data sources for supply network structures. The taxonomy describes data sources using fourteen dimensions and up to four potential characteristics. The demonstration of both research results uses a case study with two German OEMs. The case study proves the utility of the conceptual representation and of the taxonomy, as a support to identify and describe data sources about supply network structures in a systematic way.

The paper contributes to practice and research. Practitioners in charge of the data management can use the conceptual representation in a first step to identify data sources for an initial selection. In the second step, the taxonomy enables them to describe and compare the data sources, with the goal to identify the most suitable data sources for the linking process. Furthermore, they get a general understanding of the data sources, which is needed to negotiate about data source sharing and evaluate the data source portfolio. For example, potential data sources are owned by data providers or other supply chain members like supplier and logistics contractors. For researchers, the conceptual representation and the taxonomy add to the growing knowledge base of supply network visibility and are a basis for future research in the field of linking data sources.

Even with the promising results of the demonstration in a business environment, the authors cannot ensure that the taxonomy is complete. Both companies from the case studies are German automotive Original Equipment Manufacturers. Thus, more review and evaluation iterations in other industries are a valid field for future research.

References

Achilles (2017). *Helping automotive companies manage supply chain compliance and meet regulatory and legislative demands*.

AGC (2017). *Press release*. http://www.agc-glass.eu/English/Homepage/News/Press-room/Press-Detail-Page/page.aspx/979?pressitemid=3612.

Ahituv, N. (1989). "Assessing the value of information". In: *Proceedings of the tenth international conference on Information Systems - ICIS '89*. Ed. by J. I. DeGross, J. C. Henderson, and B. R. Konsynski. New York, New York, USA: ACM Press, pp. 315–325.

Awan, I., ed. (2015). *3rd International Conference on Future Internet of Things and Cloud: 24 - 26 Aug. 2015, Rome, Italy*. Piscataway, NJ: IEEE.

Basole, R. C. and M. A. Bellamy (2014). "Visual analysis of supply network risks: Insights from the electronics industry". In: *Decision Support Systems* 67, pp. 109–120.

Bergamaschi, S., D. Beneventano, F. Guerr, and M. Orsini (2011). "Data Integration". In: *Handbook of conceptual modeling*. Ed. by B. Thalheim and D. W. Embley. Berlin and New York: Springer, pp. 441–476.

Bowersox, D. J., T. Stank, and D. Closs (2000). "Ten Mega-Trends That Will Revolutionize Supply Chain Logistics". In: *Journal of Business Logistics* 21, pp. 1–16.

Bundesvereinigung Logistik e.V., ed. (2014). *7th International Scientific Symposium on Logistics: Germany, 4 - 5 Jun. 2014*.

Cagliano, R., F. Caniato, and G. Spina (2004). "Lean, Agile and traditional supply: How do they impact manufacturing performance?" In: *Journal of Purchasing and Supply Management* 10.4-5, pp. 151–164.

Chen, P. P.-S. (1976). "The entity-relationship model—toward a unified view of data". In: *ACM Transactions on Database Systems* 1.1, pp. 9–36.

Choi, T. Y. and Y. Hong (2002). "Unveiling the structure of supply networks: Case studies in Honda, Acura, and DaimlerChrysler". In: *Journal of Operations Management* 20.5, pp. 469–493.

Christopher, M. and H. Lee (2004). "Mitigating supply chain risk through improved confidence". In: *International Journal of Physical Distribution & Logistics Management* 34.5.

DeGross, J. I., J. C. Henderson, and B. R. Konsynski, eds. (1989). *Proceedings of the tenth international conference on Information Systems - ICIS '89*. New York, New York, USA: ACM Press.

eCl@ss (2017). *Classification and Product Description*. http://wiki.eclass.eu/w/index.php?title=File:ReleaseNotesBasicClassificationClass.JPG&filetimestamp=20120227160627.

Education Special Interest Group of AITP, ed. (2000). *Proceedings ISECON 2000*.

Eurostat (2017). *Your key to European statistics*. http://ec.europa.eu/eurostat/data/database.
Frohlich, M. T. and R. Westbrook (2001). “Arcs of integration: an international study of supply chain strategies”. In: *Journal of Operations Management* 19, pp. 185–200.
Grubic, T. and I.-S. Fan (2010). “Supply chain ontology: Review, analysis and synthesis”. In: *Computers in Industry* 61.8, pp. 776–786.
Harland, C. M. (1996). “Supply Chain Management, Purchasing and Supply Management, Logistics, Vertical Integration, Materials Management and Supply Chain Dynamics”. In: *Blackwell Encyclopedic dictionary of operations management* 15.
Hazen, B. T., C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer (2014). “Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications”. In: *International Journal of Production Economics* 154, pp. 72–80.
Hevner, A. R., S. T. March, and J. Park (2004). “Design Science in Information Systems Research”. In: *MIS Quarterly* 28.75-105.
ICISCE, ed. (2015). *2nd International Conference on Information Science and Control Engineering (ICISCE), 2015: 24 - 26 April 2015, Shanghai, China*. Piscataway, NJ: IEEE.
IEEM, ed. (2014). *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM): 9 - 12 Dec. 2014, Malaysia*. Piscataway, NJ: IEEE.
IHS (2014). *Auto Tech Report: Infotainment Market Overview*. http://shop.ihs.com/buy/en/ihs/auto-tech-report-infotainment-market-overview-3000006155.
IHS (2017). *Who Supplies Whom*. http://www.ihssupplierinsight.com/shop/product/261/who-supplies-whom-complete-database.
Kubler, S., M.-J. Yoo, C. Cassagnes, K. Framling, D. Kiritsis, and M. Skilton (2015). “Opportunity to Leverage Information-as-an-Asset in the IoT”. In: *3rd International Conference on Future Internet of Things and Cloud*. Ed. by I. Awan. Piscataway, NJ: IEEE, pp. 64–71.
Kulp, S. C., H. L. Lee, and E. Ofek (2004). “Manufacturer Benefits from Information Integration with Retail Customers”. In: *Management Science* 50.4, pp. 431–444.
Levary, R. R. (2000). “Better supply chains through information technology”. In: *Industrial Management* 42.3.
Leveling, J., M. Edelbrock, and B. Otto (2014). “Big data analytics for supply chain management”. In: *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. Ed. by IEEM. Piscataway, NJ: IEEE, pp. 918–922.
Li, B., L. Sun, and R. Tian (2015). “Multi-source Heterogeneous Data Exchange System for Cloud Services Platform Based on SaaS”. In: *2nd International Conference on Information Science and Control Engineering (ICISCE), 2015*. Ed. by ICISCE. Piscataway, NJ: IEEE, pp. 327–331.
Loh, T. C., S. C. L. Koh, and M. Simpson (2006). “An investigation of the value of becoming an extended enterprise”. In: *International Journal of Computer Integrated Manufacturing* 19.1, pp. 49–58.
Mukaddes, A. M. M., Choudhury, Abdul Anam Rashed, Abdul Malek, A. B. M., and J. Kaiser (2010). “Mukaddes 2010 - Developing and Information Model for Supply Chain Information Flow and its Management”. In: *International Journal of Innovation, Management and Technology* 1.2, pp. 226–231.
Murray-Rust, P., Neylon C., Pollock R., and J. Wilbanks (2010). *Panton Principles, Principles for open data in science*.

Nickerson, R. C., U. Varshney, and J. Muntermann (2013). “A method for taxonomy development and its application in information systems”. In: *European Journal of Information Systems* 22.3, pp. 336–359.

OSCM, ed. (2014). *6th International Conference on Operations and Supply Chain Management: 10-12 Dez. 2014, Bali*.

Otto, B. (2015). “Quality and Value of the Data Resource in Large Enterprises”. In: *Information Systems Management* 32.3, pp. 234–251.

Otto, B., R. Abraham, and S. Schlosser (2014). “Toward a Taxonomy of the Data Resource in the Networked Industry”. In: *7th International Scientific Symposium on Logistics*. Ed. by Bundesvereinigung Logistik e.V., pp. 382–421.

Panjivia (2017). *Powerful Insights with U.S. Customs Data*.

Patnayakuni, R., A. Rai, and N. Seth (2006). “Relational Antecedents of Information Flow Integration for Supply Chain Coordination”. In: *Journal of Management Information Systems* 23.1, pp. 13–49.

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee (2007). “A Design Science Research Methodology for Information Systems Research”. In: *Journal of Management Information Systems* 24.3, pp. 45–77.

Pidun, T. and C. Felden (2012). “Two Cases on How to Improve the Visibility of Business Process Performance”. In: *45th Hawaii International Conference on System Science*. Ed. by R. H. Sprague. Piscataway, NJ: IEEE, pp. 4396–4405.

Ramdas, K. and R. E. Spekman (2000). “Chain or Shackles: Understanding What Drives Supply-Chain Performance”. In: *Interfaces* 30.4, pp. 3–21.

Rosenzweig, E. D., A. V. Roth, and J. W. Dean (2003). “The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturers”. In: *Journal of Operations Management* 21.4, pp. 437–456.

Rozados, I. V. and B. Tjahjono (2014). “Big Data Analytics in Supply Chain Management: Trends and Related Research”. In: *6th International Conference on Operations and Supply Chain Management*. Ed. by OSCM.

Sprague, R. H., ed. (2012). *45th Hawaii International Conference on System Science: (HICSS) ; USA, 4 - 7 Jan. 2012*. Piscataway, NJ: IEEE.

Tang, C. and B. Tomlin (2008). “The power of flexibility for mitigating supply chain risks”. In: *International Journal of Production Economics* 116.1, pp. 12–27.

Thalheim, B. and D. W. Embley, eds. (2011). *Handbook of conceptual modeling: Theory, practice, and research challenges*. Berlin and New York: Springer.

Wang, X., T. N. Wong, and Z.-P. Fan (2013). “Ontology-based supply chain decision support for steel manufacturers in China”. In: *Expert Systems with Applications* 40.18, pp. 7519–7533.

Ye, Y., D. Yang, Z. Jiang, and L. Tong (2008). “Ontology-based semantic models for supply chain management”. In: *The International Journal of Advanced Manufacturing Technology* 37.11-12, pp. 1250–1260.

Yu, Z., H. Yan, and T. C. Edwin Cheng (2001). “Benefits of information sharing with supply chain partnerships”. In: *Industrial Management & Data Systems* 101.3, pp. 114–121.

Zhao, X., J. Xie, and W. J. Zhang (2002). “The impact of information sharing and ordering co–ordination on supply chain performance”. In: *Supply Chain Management: An International Journal* 7.1, pp. 24–40.

Zilbert, A. B. (2000). “A Comparative Study of Traditional Electronic Data Interchange versus Internet Electronic Data Interchange”. In: *Proceedings ISECON 2000*. Ed. by Education Special Interest Group of AITP.