

## From Customer Data to Smart Customer Data: The Smart Data Transformation Process

*Thang Le Dinh<sup>1,\*</sup>, Nguyen Anh Khoa Dam<sup>1,2</sup>, Chan Nam Nguyen<sup>1,3,4</sup>,  
Thi My Hang Vu<sup>1,3,4</sup>, and Nguyen Cuong Pham<sup>3,4</sup>*

<sup>1</sup> Business School, Université du Québec à Trois-Rivières, Canada

<sup>2</sup> The University of Danang, University of Science and Technology, Vietnam

<sup>3</sup> Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam

<sup>4</sup> Vietnam National University, Ho Chi Minh City, Vietnam

**Abstract.** Nowadays, smart data has emerged as a new trend in creating more business value for enterprises that is defined as the data that is gathered and processed to create new insights to support business decisions. However, the transformation from data into actionable insights is still a real challenge for enterprises. For this reason, this paper presents a smart data transformation process, which aims at transforming customer data into smart customer data in order to offer actionable insights. The purpose of the study is to propose a transformation process that can be used to operate a knowledge structure for a smart service system, which can manage and deliver smart data as a service. The process covers the three dimensions of a service system: Data processing corresponding to the engineering dimension, information processing corresponding to the science dimension, and knowledge processing corresponding to the management dimension for knowledge processing. Accordingly, a case study on the smart data transformation process of a customer journey management system as a smart service system is presented to demonstrate the application of the proposed process.

### 1 Introduction

Nowadays, smart data is emerging as a new source for service innovation [1], which is defined as the data gathered and processed to create new insights to support business decisions. Smart data has emerged as a trend in creating more business value for enterprises. However, the link between data and business value to generate actionable insights is still missing, leading to a challenge for the information system research community [2]. Accordingly, there is a need for a smart service system to handle the transformation from data into actionable insights. A smart service system is a typical service system, which is “capable of learning, dynamic adaptation, and decision-making based upon data received, transmitted, and/or processed to improve its response to a future situation” [3].

Focusing on the operations of the smart service system for supporting smart customer data, the paper proposes a smart data transformation process that can be used for operating a knowledge structure for a smart service system to promote customer intelligence [4]. In other words, the paper presents a smart data transformation process for transforming customer data into smart customer data in order to offer actionable insights and to deliver smart data as a service.

The paper is structured as follows. Section 2 presents the theoretical background. Section 3 explains the research design and Section 4 continues with the smart data transformation process. Accordingly, Section 5 introduces a case study about the smart data transformation process for a particular smart service system: a customer journey

---

\* Corresponding author: [thang.ledinh@uqtr.ca](mailto:thang.ledinh@uqtr.ca)

management system to demonstrate the application of the proposed process. Section 6 ends the paper with conclusions and future research directions.

## 2 Theoretical backgrounds

### 2.1 Smart customer data

Smart customer data can serve as a great means for generating actionable insights and for providing smart services to customers [4]. It is contended that smart customer data is advantageous; however, what actions enterprises can make upon these data are still debating issues [5]. Regardless of the significantly increased amount of customer data, enterprises seem to lose track in defining what smart customer data are and how to make optimal decisions [6]. Literature defines smart customer data in the sense that they can produce actionable insights for data-driven decisions [7,8].

### 2.2 Applications of smart customer data

Considering the importance of smart customer data for smart services, this part continues exploring the literature to clarify the different applications of smart customer data.

#### Promoting smart service development and innovation.

The application of smart customer data in smart service development would yield profitable outcomes [9]. Actionable insights from smart customer data keep track of changes in markets and customer preferences as they are extracted from customer interactions on digital platforms [8]. Nowadays, customers express their voices on social media, which creates a significant amount of real-time user-generated data for the development of smart services [5,7]. Analyzing and applying these real-time data would reveal deep insights on customers and offer smart service relevant to the context of customers [6]. For example, smart customer data are applied in developing context-aware recommender systems to make relevant recommendations.

#### Optimizing customer journeys.

Smart customer data can be applied to design and optimize customer journeys, which consist of *recognizing* customer needs, *requesting* a product or service that might meet their needs, and *responding* to the delivery of service and products [10]. In other words, the three stages of recognizing, requesting, and responding represent the pre-purchase, purchase, and post-purchase of a customer journey [11]. Accordingly, service providers attempt to map customer journeys with key touchpoints in each stage [12]. Identifying touchpoints is significant as they are interactive opportunities for potential customers to engage with service providers [13].

#### Facilitating customer relationship management.

To build long-term relationships with customers, it is important to offer products and services that meet customers' needs [4]. Smart customer data empowers enterprises with customer orientation in which enterprises consider customer preferences for products/service offerings. It is noted that smart customer data also promotes product and service orientation with innovative offerings to attract and retain customers [14]. In addition, smart customer data equips enterprises with competitor orientation [6]. Therefore, enterprises will be able to keep track of competitors' activities, anticipate competitors' strategies, and learn from competitors' products. Customer orientation coupled with product/service orientation and competitor orientation sets a solid foundation for smart customer data to support enterprises in managing customer relationships.

#### Improving data-driven decisions.

The application of smart customer data empowers enterprises to boost financial performance and achieve substantial profit gains [1]. In general, enterprises that take advantage of smart customer data can outperform their competitors with 6% more on profits [15]. Top-performing enterprises tend to apply customer intelligence to improve data-driven decisions [16,17]. Furthermore, customer intelligence, which promotes the transformation of information technology, is promised to be a strategic economic resource for enterprises.

#### Improving customer loyalty and lifetime value.

Smart customer data would be the optimal solution for enterprises to gain deep insights into customer behaviors, preferences, and trends. Thanks to smart customer data, enterprises can satisfy customers' needs and improve their loyalty. Customer loyalty, which indicates the commitment and intention of customers for repurchases, relies on the satisfaction with the core product and service as well as the delivery process. Products and services derived from customer intelligence are proved to induce customer loyalty due to the high level of customer satisfaction. The more

satisfied customers are, the more likely they are expected to spend. This indicates the greater lifetime value that customers can offer for enterprises [18,19].

### 3 Research design

#### 3.1 Problem statement

Actionable insights from customer data are meaningful knowledge that can be acquired from the application of data analytic techniques [20]. In fact, it is a real challenge for moving forward to insights if there is a lack of a background to understand why it is important in some business situations, which require context-related data [3].

Smart data is a good start with reliable and verified data; those data need support to become enriched, contextualized, and featured in order to generate actionable insights [21]. However, the current studies often focus on a specific level of data transformation (i.e., data, information, or knowledge level). Therefore, there is an urgent need for frameworks, models, and processes for transforming raw data into smart data in a coherent manner.

#### 3.2 Research question and objectives

This study aims at proposing a *smart data transformation* process; hereafter called the SDT process, based on the service science perspective to promote customer intelligence. The proposed process includes three levels: data processing, information processing, and knowledge processing. According to the perspective of service science, the data processing level corresponds to the engineering dimension, the information processing corresponds to the science dimension, and the knowledge processing level corresponds to the management dimension [22].

The key research question is: “*How to support the transformation of customer data into smart data in a smart service system*”.

The previous section presents the literature review to better understand the potentials of smart customer data for generating actionable insights and the contributions of the proposed study. At first, actionable insights fill the gap between smart customer data and business values [5]. Therefore, actionable insights involve practical applications rather than theoretical contributions or mere reports [2,20].

With an aim to demystify actionable insights, the SMART characteristics of the SDT process, including **Strategy-oriented**, **Measurable**, **Analytic-based**, **Result-visualized**, and **Transforming** are presented as follows:

- *Strategy-oriented*: actionable insights should be aligned with the goals and strategies of an enterprise to drive actions [2,23].
- *Measurable*: actionable insights should be measurable by identifying relevant data for key performance indicators related to the strategic goals [7].
- *Analytic-based*: actionable insights result from the applications of descriptive, predictive, and prescriptive analytics to transform data into insights [7,24].
- *Result-visualized*: actionable insights need to be visualized through digital dashboards, graphs, or models to support the decision-making process [7].
- *Transforming*: actionable insights should transform businesses in terms of i) Economic value – how to make the most profits from customers), ii) Social value – how to use customers for social influence, and iii) Cognitive value – how to take advantage of customers’ knowledge and experience [25].

#### 3.3 Research methodology

In this study, different processes to transform data and information into knowledge and insights will be examined to determine the specific requirements of the transformation process. Thus, the SDT process as the key deliverable of the study is presented based on the design science research methodology (DSR) [26].

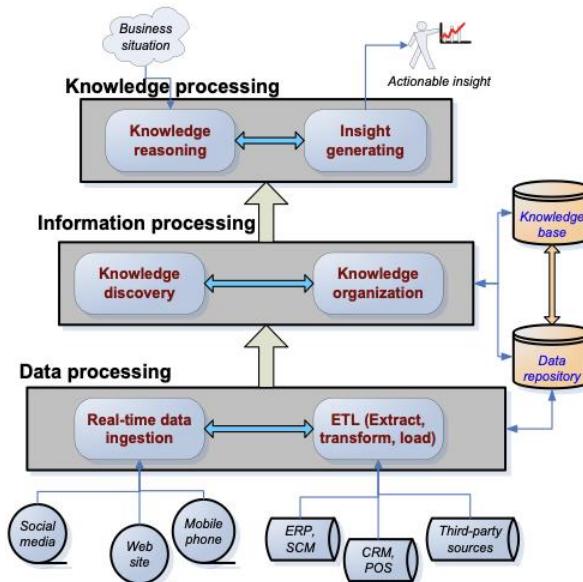
Accordingly, in terms of conception and design, this study proposes three types of artefacts: constructs, models, and processes [26]. *Constructs* are objects and modules involving three phases of the SDT process (Data processing, Information processing, and Knowledge processing) (cf. Fig. 1). For instance, key constructs of the information processing phase are two modules (Knowledge Organization, Knowledge Discovery) and a knowledge base. *Models*, which present interrelationships between constructs [26], primarily involve abstract structures for organizing databases and knowledge bases. *Processes*, which mostly focus on methods and/or steps to produce constructs, include

the SDT process, ETL and real-time data processing (cf. Section 4.1), and knowledge base construction process (cf. Section 4.2).

In terms of demonstration and evaluation of the DSR methodology, a real-world application (i.e., a customer journey management system as a smart service system) of the SDT process as an illustration, including the technical experiment and case study, is illustrated in Section 5.

## 4 Smart data transformation process

The smart data transformation process is introduced in Fig. 1. which includes three processing levels: Data processing, information processing, and knowledge processing. Each level may have different components, including *constructs* and *methods* of the process [27]. This section continues with data processing, information processing, and knowledge processing. Since the main contribution of the study is smart data transformation, the knowledge processing is presented in more detail.



**Fig. 1.** Smart data transformation process.

### 4.1 Data processing

The term "data processing" refers to the transformation of data from a range of various low-level data sources into purposed data that can then be stored in specified data storages, called data repositories, which are most notably database management systems [22]. In the era of big data, customer data is increasingly arriving from a variety of data sources that are too dissimilar in the format, including relational databases, websites, mobile devices, company servers, social media, and third-party data providers [20]. By leveraging the development of new technologies, particularly big data tools, it is possible to capture and store a diverse collection of customer data in a data repository, which is constantly available for high-level tasks for generating customer smart data.

In this study, data processing consists of the following components: *Data source*, *Data repository*, *Real-time data ingestion*, and *ETL tool*. These components enable the processing of a variety of customer data sources in order to support data collection, provision, and analytics [28].

**Data source.**

Currently, customer data is derived from a variety of data sources that fall into two main categories:

- *Batch data sources*: store large amounts of data in bulk, such as customer relationship management systems (CRM), Enterprise Resource Planning (ERP), SCM (Supply Chain Management), POS (Point of Sale), and third-party services [29]. Customer data from these sources can be in the form of customer profiles or transactions, and it is typically stored in SQL databases, NoSQL databases, or log files [30,31]. Because of the size of this data, it is scheduled to generate daily customer metric reports, which take a long time to process and calculate [32].
- *Real-time data sources*: sources that customer data can be created continually every millisecond as a result of customers engaging with websites, mobile devices, or social media platforms [33]. The majority of the data from these sources is unstructured or semi-structured and comprises the following: message, email, comment, post, like, image, and clickstream [34]. With these types of data, it is critical that the system processes and analyses them within milliseconds or seconds to ensure that customers receive timely responses or services in order to improve their experience [19,35].

**Real-time data ingestion.**

This component enables the collecting, processing, and temporary storage of all sorts of customer data generated in real-time from social networks, websites, and mobile devices in order to ensure that data is sent to the next level of the system without interruption [36]. The most widely used real-time data processing technologies are REST APIs, Kafka, RabbitMQ, and ActiveMQ [37]. Following this stage, the data can be processed in real-time for streaming analysis or stored in a data repository.

**ETL tool.**

This component facilitates the processing of significant batches of data by reading data from source files or databases and then converting it to the required format for storing in data repositories (Database Management System or Hadoop Distributed File System) [28]. This process may contain several sub-data pre-processes, including data cleansing, normalization, deduplication, structure mapping, and sub-data processes, including validating, sorting, aggregating [38]. Numerous popular ETL (Extract, Transform, and Load) solutions are currently available for usage at this step to bulk load data from a range of data sources, including CSV files, NoSQL databases, and relational databases, into a particular database management system. Apache Sqoop is commonly used to export and import huge volumes of data effectively between the Hadoop Distributed File System (HDFS) and structured data storage such as relational databases [39].

**Data repository.**

Data repository is defined as a storage component that manages a huge volume of customer data processed from ETL tools and real-time data ingestion [33]. From the perspective of customer focus, customer data can be categorized into 4 main groups [29]:

- *Demographic data* provides information about groups of customers based on their age, gender, residence, family status, or income [20,40]. Customer relationship management (CRM) systems, customer data platforms (CDP), thirty-party data providers, and social media are the key sources of this sort of data.
- *Behavioral data* includes data generated by a customer's interaction with a business service or company product [19,41]. This type of data can comprise critical customer behaviors (such as registration, login, add-to-cart, purchase, page view, and clickstream), as well as consumer habits (communication channel, device, browser, and operating system) [4]. Websites, mobile applications, CRM systems, marketing automation systems, call centers, help desks, and billing systems are all common sources of behavioral data.
- *Transactional data* includes information about customer purchases such as orders, total amount, kind of item, the number of items, transaction date, payment methods, and shipping addresses [41,42]. This sort of data can be stored in an organization's RDMS or through third-party services in the form of transaction, order, billing, and invoice records.
- *Psychographic data* includes information about customer emotions, evaluation, preferences regarding the company's services and products [8]. Customer comments, likes, reviews, and ratings on social media, as well as customer feedback from online surveys, can all be examples of psychographic data.

Recently, numerous prominent technologies have been employed for big data repositories. Relational database management solutions can be used with moderate amounts of data. Hadoop Distributed File System (HDFS), on the

other hand, is a data repository suitable for storing large amounts of data because it is a well-known fault-tolerant distributed file system of Apache Hadoop, a dominant framework for big data processing with large infrastructures being deployed and used in a variety of application fields [43].

## 4.2 Information processing

The information processing level aims at transforming pre-processed data from the data processing level into well-structured information to assist business insight generation [28]. Turning data into useful information is a challenging task due to the complexity and diversity of the data.

The study considers herein two significant issues: i) How to provide support for knowledge discovery and ii) How to structure data in an appropriate organization for knowledge retrieval. Accordingly, the information processing level consists of the following components: *Knowledge discovery*, *Knowledge base construction*, and *Knowledge base*.

### **Knowledge discovery.**

Knowledge discovery activities involve transforming data, primarily structured in the tabular form, into the knowledge base by applying matching business analytics techniques. Those activities may also involve extracting knowledge entities and relations from unstructured or semi-structured data by applying data mining and/or machine learning techniques.

### **Knowledge base construction.**

The knowledge base (KB) method aims at structuring a large amount of heterogeneous data from diverse sources in a suitable organization. There are two well-known approaches for this construction: top-down and bottom-up [44]. The first approach consists of pre-defining a knowledge model, then importing knowledge instances into the model. The second one, on the other hand, focuses on knowledge extraction from raw data. This study proposes a hybrid approach that leverages the pre-defined knowledge component (i.e., CAK model [45]) and seems suitable to the research problems concerned. According to the adopted method, this study proposes the KB construction component consisting of four essential elements (know-what, know-who, know-how, know-why).

### **Knowledge base.**

Knowledge base, which is a storage component, is serving the reasoning process, is built on top of the abstract structure through knowledge importing and extraction activities. Based on the CAK model [45], an abstract structure for KB is proposed to feature knowledge components: know-what, know-who, know-how, know-why, know-with, know-when, and know-where [1-3] and their links to different data in the data repository.

## 4.3 Knowledge processing

Knowledge processing is defined as the process to put knowledge through the forms of actions. To put it another way, knowledge processing is the process of making values from the knowledge components for actionable insights [8,46]. The knowledge processing level includes two methods: *Knowledge reasoning* and *Insight generating*.

### **Knowledge Reasoning.**

Knowledge reasoning provides services to exploit the knowledge base. For knowledge reasoning, various methods have been studied in recent years [44], e.g., reasoning via patterns, query languages, inference rules, and so forth. In this study, the reasoning engine, whose purpose is to serve the higher levels, will be implemented as APIs relying on a knowledge query language (e.g., SPARQL, Cypher, etc.) and a set of inference rules. According to the adopted method, a service receives a knowledge component (e.g., know-what, know-who [29]) as an input, then exploits the KB to retrieve all direct and/or indirect knowledge components related to the input considered. In particular, knowledge reasoning enables different services featuring various types of knowledge. More precisely, this study focuses primarily on the interrelations among customer (know-who), product/service (know-what), process/transition (know-how), and customer value (know-why). A set of predefined patterns is proposed to determine the reasoning process based on the input knowledge component (Table 1). The output of the reasoning process can be displayed by various tools (e.g., dashboard, graph, or simple tabular structure).

**Table 1.** Knowledge reasoning patterns.

Pattern	Input	Output	Description
Know-what based pattern	Behavioral data Psychological data	Product/Service Profiles	Exploiting customer behaviors and preferences to identify the most suitable products/services for customers' needs.
Know-who-based pattern	Demographic data Transactional data	Customer Profiles Customer Clusters	Exploiting a customer's historical data, especially transactional data, to discover customer profiles (i.e., What did they buy? When? Where?) to support customer clustering based on customers' similar characteristics.
Know-how-based pattern	Behavioral data	Activities/Interactions, coming up with related products/services and involved customers	Exploiting the way how customers interact with products/services and/or other customers.
Know-why-based pattern	Behavioral data Psychological data Transactional data	Interrelationship between different knowledge components, coming up with context information (location, time)	Exploiting customer behaviors, preferences, and historical purchases to discover interesting relationships between know-who, know-with, and know-how components, which can help to generate business values.

### Insight generating.

From the perspective of smart customer data, the paper identifies different types of actionable insights and applies relevant knowledge components (know-what, know-how, know-why, and know-who) through each stage of the customer management process, including customer profiling, customer engagement, customer experience, and customer value [4,41].

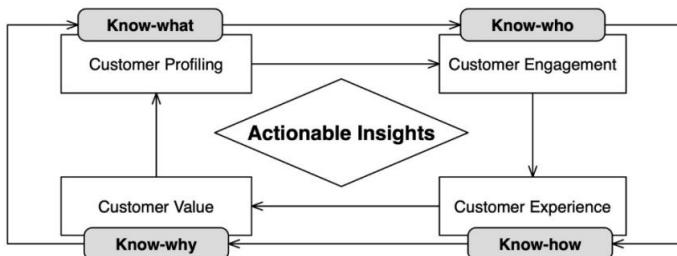
**Fig. 2.** Insight generating.

Fig. 2. illustrates the knowledge component corresponding to each stage of the customer management process for generating actionable insights: Customer profiling, customer engagement, customer experience process, and customer value process.

**Customer profiling** focuses on the *know-what* component by developing products/services that can meet customers' needs [11,25]. Derived from behavioral and psychological data, the *know-what* component reveals the preferences and behaviors of customers [47]. With an aim to promote actionable insights as a service, the *know-what* component can be demonstrated through graphs of product/service profiles or developed into product/service ontology [35]. In this light, marketers can rely on such actionable insights for product development and innovation. Accordingly, these insights deal with product/service innovation in optimizing product/service features and characteristics that offer value for customers [4].

**Customer engagement** relies on the *know-who* component to promote interactions with customers. Enterprises make use of *know-who* to develop customer profiles by identifying similar historical purchases from demographic and transactional data [48,49]. The *know-who* component contains knowledge on demography (age, gender), buying behaviors (needs, purchasing power, preferences, lifestyle), and purchasing attributes (recency, frequency, size)

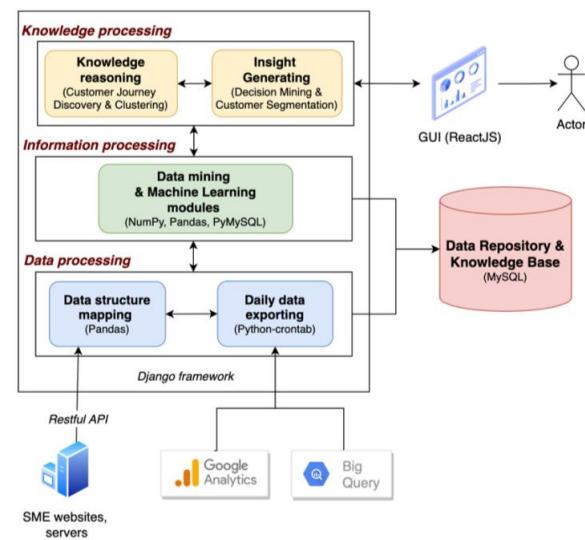
[19,50]. Actionable insights acquired from *know-who* can be illustrated as customer profiles on a dashboard. Such dashboards would facilitate the process of customer segmentation, customer engagement, and customer profile management. With the support of actionable insights from *know-who*, enterprises will be able to define the target audience and choose the most profitable segment. Furthermore, enterprises will be able to better adjust their marketing strategies, particularly strategies related to segmentation, targeting, and positioning for optimal outcomes [4].

**The customer experience process** highlights the importance of the *know-how* component to manage customer experience and reinvent customer journeys [11,41]. Derived from behavioral data, the *know-how* component contains knowledge related to the interactions of customers-products/services and customers-customers [46,51]. From the perspective of knowledge processing, actionable insights from *know-how* can be developed into dashboards or interactive reports so that marketers can keep track of customer experience in each journey from pre-purchase and purchase to post-purchase [52]. With the support of such dashboards, marketers can identify key touchpoints to acquire customers and improve sales [16,20].

**The customer value process** aims at maximizing customer values for enterprises through the transformation of *know-what* into *know-why* – what to do next with customers to achieve the enterprise's goals [53,29]. The *know-why* component contains actionable insights related to i) Economic value – how to make the most profits from customers), ii) Social value – how to use customers for social influence, and iii) Cognitive value – how to take advantage of customers' knowledge and experience [25,54]. In terms of economic value, the *know-why* component contains actionable insights related to customer lifetime value and up/cross-selling. On the other hand, actionable insights on social values take advantage of customers' social status and networks to impact other customers through word of mouth (WOM) [8,25]. Finally, actionable insights on cognitive value figure out what value to offer for customers, what resources to provide, and what engagement forms to facilitate customer co-creation [55,56].

## 5 A customer journey management system

This section of the paper continues with a case study to demonstrate how the SDT process is used for designing and developing a customer journey management system, called *Customer Journey Master* (CJMA). CJMA aims at assisting small and medium-sized enterprises (SMEs) in generating customer journey insights. The three processes (Data, Information, and Knowledge processing) of CJMA are presented in the following parts (Fig. 3.).



**Fig. 3.** Customer journey management system.

Django framework [57], a free and open-source Python web platform, serves as the backbone for the CJMA system. Smart data transformation phases are implemented on the Django framework by integrating Python libraries, which are used for data processing, data analytics, data science, and machine learning. At the Data processing level, Daily data exporting and Data structure mapping are the two main components that help in collecting, extracting, and pre-processing raw data into structured data before storing it in data repositories. With the help of the Python-crontab package, Daily data exporting automates the collection of customer data from Google Analytics and Big Query, which exports daily customer events and demographic data into large volume csv files. These file data, as well as tracking data collected from the SME's website and server via the predefined API, would be cleaned, normalised, validated, and then restructured according to CJMA's data repositories by using Panda's library. MySQL was utilized for data storage because it is an open-source relational database management system that is suitable for SMEs [58]. The system provides data mining and machine learning modules at the Information processing level to assist in the transformation of processed data into suitable information for analysis. These modules are built on a variety of Python libraries, including NumPy (for data vectorization) [59], Pandas (for tabular format transformation) [60], and PyMySQL (for MySQL database connection) [61]. Finally, at the level of Knowledge processing, several reasoning and insight generating components are implemented for customer journey analytics such as metric reports, customer segmentation, customer journey visualisation, customer journey clustering, and customer decision mining. In this stage, the Scikit-learn library (which provides various classification, clustering, and decision tree algorithms) [62] and PM4Py library (which provides methods of process discovery and sequential mining) [63] are used to extract useful customer knowledge and insight. To gain actionable insights, business actors would interact with the CJMA system via a graphical user interface built with ReactJS. Dashboards, scorecards, process models, report tables, and bar charts are used to display actionable insights.

## 5.1 Data processing

The CJMA data sources include customer data from CRM systems, third-party platforms such as Google Analytics, company websites, and company mobile applications. Google Analytics is a great source of demographic and behavioral data since it can track customer actions like clickstreams and events that customers initiate on e-commerce websites that use Google Analytics tracking code. These data comprise not only customer information such as customer activity history, but also customer preferences such as communication channels, devices, browsers, and operating system information. Company websites and mobile applications can also capture behavioral and transactional data through consumer interactions.

device_browser	device_os	None	device_category	geo_continent
browser	operatingSystem	customer_id action_type visit_time channel_type device_browser device_os device_category geo_continent geo_country geo_city source_name source_url ad_url ad_content campaign_url campaign_content store_id store_address employee_id	deviceCategory	continent
Safari	iOS		mobile	Americas
Safari	iOS		mobile	Americas
Safari	iOS		mobile	Americas
Safari	iOS		mobile	Americas
Safari	iOS		mobile	Americas

**Fig. 4.** CSV file importing and structure mapping.

In terms of the CJMA model, the data processing level leverages several data sources and data processing procedures to extract, transform, and load data into a centralized data repository. Concerning the ETL process, the CJMA system has a data structure mapping mechanism that facilitates the integration of customer data from huge csv files. For instance, customer journey data from Google Analytics, containing several rows of customer events can be daily exported into csv files (Fig. 4.) and then be loaded into CJMA data storage to combine and create daily customer

metric reports. In terms of real-time data ingestion, CJMA offers a list of predefined Restful APIs that other systems, such as websites and mobile applications, can utilize for sending real-time customer event data to the system.

MySQL was chosen as the data repository for CJMA due to the system's target audience of SMEs with limited budgets for infrastructure development. The data repository contains several primary data tables relating to the four types of customer data mentioned previously: demographic (Customer), behavioral (Touchpoint, Action, Device Category, OS, Browser), psychographic (Email, Post, Comment, Rating), and transactional (Order). Customer data in the database management systems (DBMS) can now be exported into csv/json files, and then be accessible to higher layers through APIs to acquire new customer knowledge.

## 5.2 Information processing

In the case of CJMA running example, data comes from two primary sources: traditional database systems and Google Analytics. After the pre-processing process, the data is transferred, validated, and reorganized in the data repository to feature data classification and indexation based on the corresponding knowledge components in the knowledge base. Thus, those data are promptly forwarded to real-time processing for analysis in order to discover new knowledge such as demographic data, behavioral data, transactional data, and psychographic data.

**Demographic data**, coming from enterprise systems and Google Analytics, may be used to exploit the similarity among customers (know-who) based on their profiles. This information is essential to segment customers, better understand the target group of customers, create personas, and plan advertising campaigns [41].

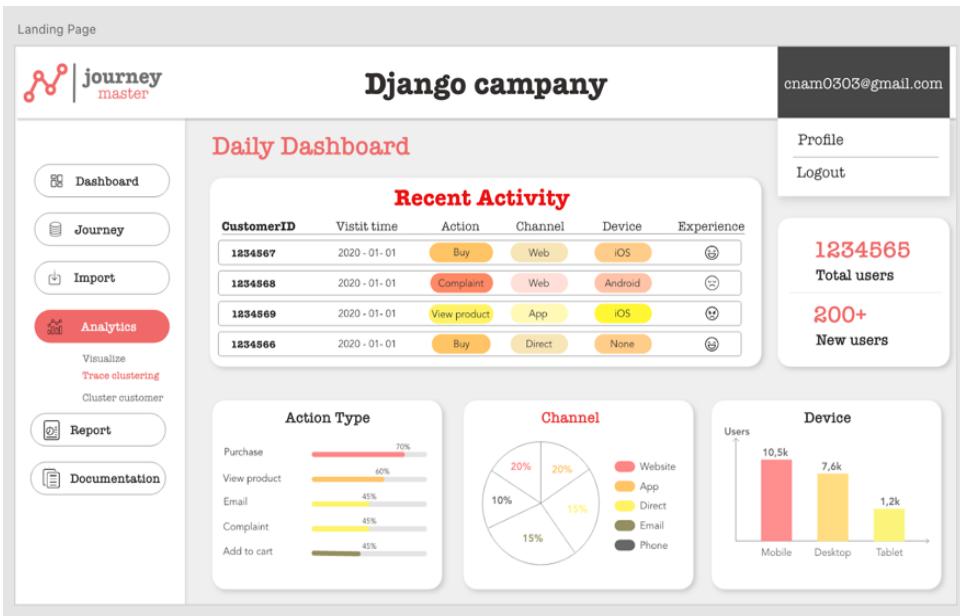
**Behavioral data**, mainly containing customers' clickstream and event information, may be used to exploit the interrelations among customers (know-who) and products (know-who) through their interactions (know-how) on enterprise websites, as well as context information (i.e., where and when customers' interactions take place). This information assists to analyze the customer journey, personalize customer acquisition, and forecast consumer behaviors in order to enhance the customer experience [18,64].

**Transactional data**, retrieved from customers' orders, may be used for daily revenue reports, consumer behavior analysis, purchase likelihood prediction, frequent itemset, and sequential pattern mining [65].

**Psychographic data**, primarily involving comments and ratings of a customer (know-who) for a product (know-what) via his/her interactions (know-how), may be used to discover customer preferences. This information is frequently utilized for text mining, emotion classification, and sentiment analysis in order to turn raw data (text) into customer smart data [66].

## 5.3 Knowledge processing

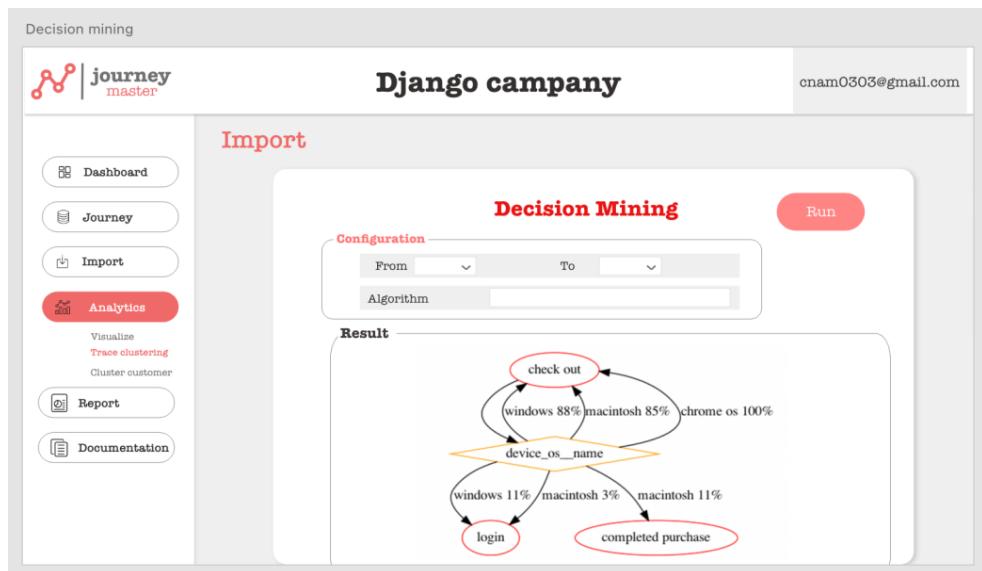
The CJMA not only supports the knowledge and insights as mentioned above but also offers some additional insights related to customer journeys such as customer journey discovering, customer journey clustering, and customer decision mining.



**Fig. 5.** Knowledge processing interface.

The Knowledge reasoning component focuses on customer journey discovering and clustering:

- *Customer journey discovery* is a function that maps an event log to a process model, which can fully describe a customer journey [67]. The basic idea of journey discovery is as the following: given an event log  $L$ , process discovery aims at mining the workflow of customers, in terms of actions and behaviors (Fig. 5.). Using the discovered process model, users can determine the main flow of the customers when they interact through the different touchpoints, such as websites or mobile apps, and then build a recommendation system or improve the available services.
- *Customer journey clustering*, also called Discovering behavior pattern, aims at clustering the event log into several small event logs based on their similar properties to produce multiple simple process models. In other words, the CJMA attempts to map customer journeys with key touchpoints in each stage [20]. Thus, users can visualize the process model as a graph, which can be used to predict the behavior of customers in order to improve customer experience, recommend appropriate products, and create a better marketing strategy [40].



**Fig. 6.** Customer decision mining.

The Insight generating component deals with *customer decision mining*, which is an extended version of process enhancement to mine the workflow perspective of behaviors. Fig. 6. illustrates a decision process model, which is a predicted customer journey including future activities which customers can decide to take based on information about the customer's device usage habits (device category, operating system, etc.). Indeed, smart customer data can be applied to optimize customer journeys such as recognizing customer needs, requesting a product or service meeting their needs, and responding to the delivery of service and products [18].

## 6 Conclusion

Nowadays, smart data such as smart customer data becomes a new source for generating actionable insights to create more business value for enterprises. In order to establish the solid link between data and business value, the paper presents a *smart data transformation process*, called the SDT process, for transforming customer data into smart customer data to deliver smart data as a service and to promote customer intelligence. The SDT process is presented based on the perspective of service science, including the data, information, and knowledge processing levels.

In order to develop the SDT process, different data transformation processes have been investigated. Firstly, the general processes such as the Data-Information-Knowledge-Wisdom (DIKW) process [68] and the knowledge management process are evaluated [69]. Secondly, a specific process at the data level: Extraction-Transformation-Loading (ETL) process is also explored such as [70]. Finally, several data mining processes at the information level such as KDD, CRISP-DM, and SEMMA [71] are examined.

The main differences between the SDT process and other processes are that the SDT process supports the data transformation in a coherent manner and at all the levels of the transformation. Compared with the general processes such as the knowledge management process and DIKW, the SDT process focuses on processing mostly structured data to create actionable insights, whereas the other processes focus on different types of data. Compared with the other specific processes, the SDT process deals with all three levels (data, information, and knowledge), whereas the ETL concerns the data level, the data mining processes concern mostly the information level. In general, similar processes have paid little attention to Real-time data ingestion and Insight generating. It is argued that the SDT process is one of the first approaches that focus on smart data transformation based on the service science perspective.

Concerning the implications of our work in practice, the SDT process can be enhanced and customized to meet specific requirements of the smart service system of an enterprise in order to effectively and efficiently offer smart

services. Concerning the implications for research, this approach needs to experiment on a broader scale with different types of smart services for promoting customer intelligence.

Finally, our future studies aim at enhancing the approach for context-aware smart services to support artificial intelligence (AI) based services such as chatbot, recommender systems, and insight engines. Moreover, the SDT process can be also extended so that smart data can support the new trends of today's business ecosystem such as resilience and sustainability [72].

## References

1. T. Le Dinh, N. A. K. Dam, Smart data as a service. *Proceedings of the International conference on exploring service science (IESS) 2.1*, (2021)
2. B. Dykes, Actionable insights: The missing link between data and business value. *Forbes* (2016)
3. A. Medina-Borja, Editorial Column—Smart Things as Service Providers: A Call for Convergence of Disciplines to Build a Research Agenda for the Service Systems of the Future. *Service Science* **7**(1), ii-v (2015)
4. N. A. K. Dam, T. Le Dinh, W. Menvielle, Towards a Conceptual Framework for Customer Intelligence in the Era of Big Data. *International Journal of Intelligent Information Technologies (IJIIT)* **17**(4), 1-17 (2021)
5. J. Shim, R. Taylor, Purchase-Based Analytics and Big Data for Actionable Insights. *IT Professional* **21**(5), 48-56 (2019)
6. V. S. Sharma, R. Mehra, S. Podder, A. P. Burden, A journey towards providing intelligence and actionable insights to development teams in software delivery. *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, (IEEE, 2019)
7. I. Triguero, D. García-Gil, J. Maillo, J. Luengo, S. García, F. Herrera, Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9**(2), e1289 (2019)
8. A. Wieneke, C. Lehrer, Generating and exploiting customer insights from social media data. *Electronic Markets* **26**(3), 245-268 (2016)
9. J. Hester, T. R. Miller, J. Gregory, R. Kirchain, Actionable insights with less data: guiding early building design decisions with streamlined probabilistic life cycle assessment. *The International Journal of Life Cycle Assessment* **23**(10), 1903-1915 (2018)
10. N. Siggelkow, C. Terwiesch, The Age of Continuous Connection. *Harvard Business Review* (2019)
11. A. Rawson, E. Duncan, C. Jones, The Truth About Customer Experience. *Harvard Business Review* (2013)
12. A. Stein, B. Ramaseshan, Towards the identification of customer experience touch point elements. *Journal of Retailing and Consumer Services* **30**, 8-19 (2016)
13. K. N. Lemon, P. C. Verhoef, Understanding customer experience throughout the customer journey. *Journal of marketing* **80**(6), 69-96 (2016)
14. F. Colbert, D. C. Dantas, Customer Relationships in Arts Marketing: A Review of Key Dimensions in Delivery by Artistic and Cultural Organizations. *International Journal of Arts Management* **21**(2) (2019)
15. A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, D. Barton, Big data: the management revolution. *Harvard business review* **90**(10), 60-68 (2012)
16. U. Sivarajah, M. M. Kamal, Z. Irani, V. Weerakkody, Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research* **70**, 263-286 (2017)
17. P. Zerbino, D. Aloini, R. Dulmin, V. Mininno, Big Data-enabled customer relationship management: A holistic approach. *Information Processing & Management* **54**(5), 818-846 (2018)
18. E. W. T. Ngai, L. Xiu, D. C. K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications* **36**(2), 2592-2602 (2009)
19. Y. Yan, C. Huang, Q. Wang, B. Hu, Data mining of customer choice behavior in internet of things within relationship network. *International Journal of Information Management* **50**, 566-574 (2020)
20. S. Gupta, A. Leszkiewicz, V. Kumar, T. Bijmolt, D. Potapov, Digital analytics: Modeling for insights and new methods. *Journal of Interactive Marketing* **51**, 26-43 (2020)
21. K. A. Whittler, Stop Focusing On Big Data And Start Focusing On Smart Data. *Forbes* (2019)
22. T. Le Dinh, T. T. T. Pham, Information-driven framework for collaborative business service modelling. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)* **3**(1), 1-18 (2012)
23. M. Murphy, J. Barton, From a Sea of Data to Actionable Insights. (2014)
24. A. Levenson, A. Fink, Human capital analytics: too much data and analysis, not enough models and business insights. *Journal of Organizational Effectiveness: People and Performance* (2017)

25. K. Xie, Y. Wu, J. Xiao, Q. Hu, Value co-creation between firms and customers: The role of big data-based cooperative assets. *Information Management Science* **53**(8), 1034-1048 (2016)
26. A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research. *MIS Quarterly* **28**(1), 75-105 (2004)
27. A. Hevner, S. Chatterjee: Design science research in information systems. In: *Design research in information systems*. pp. 9-22. Springer, (2010)
28. T. Le Dinh, T.-C. Phan, T. Bui, M. C. Vu, Towards a Service-Oriented Architecture for Knowledge Management in Big Data Era. *International Journal of Intelligent Information Technologies* **14**(4), 24-38 (2018)
29. T. L. Dinh, N. A. K. Dam, Towards Smart Customer Knowledge Management Systems. *International Conference on Computational Collective Intelligence*, (Springer, 2021)
30. K. Li, V. Deolalikar, N. Pradhan, Big data gathering and mining pipelines for CRM using open-source. *2015 IEEE International Conference on Big Data (Big Data)*, (IEEE, 2015)
31. G. Eken, A. Erkollar: The potential and capabilities of NoSQL databases for ERP systems. In: *Advanced MIS and Digital Transformation for Increased Creativity and Innovation in Business*. pp. 147-168. IGI Global, (2020)
32. M. Kiran, P. Murphy, I. Monga, J. Dugan, S. S. Baveja, Lambda architecture for cost-effective batch and speed big data processing. *2015 IEEE International Conference on Big Data (Big Data)*, (IEEE, 2015)
33. J. O. Chan, Big data customer knowledge management. *Communications of the IIMA* **14**(3), 5 (2014)
34. J. Abawajy, Comprehensive analysis of big data variety landscape. *International journal of parallel, emergent and distributed systems* **30**(1), 5-14 (2015)
35. M. Alnoukari, A framework for big data integration within the strategic management process based on a balanced scorecard methodology. *Journal of Intelligence Studies in Business* **1**(1) (2021)
36. G. Pal, G. Li, K. Atkinson, Big data real time ingestion and machine learning. *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, (IEEE, 2018)
37. P. Singh, B. Chaitra, Comprehensive Review of Stream Processing Tools. (2020)
38. C. French, Data processing and information technology. Cengage Learning EMEA, (1996)
39. S. Pippal, S. P. Singh, D. S. Kushwaha, Data Transfer From MySQL To Hadoop: Implementers' Perspective. *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies*, (2014)
40. N. A. K. Dam, T. Le Dinh, W. Menvielle, Marketing Intelligence From Data Mining Perspective – A Literature Review. *International Journal of Innovation, Management and Technology* (2019)
41. M. Holmlund, Y. Van Vaerenbergh, R. Ciuchita, A. Ravald, P. Sarantopoulos, F. V. Ordenes, M. Zaki, Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research* (2020)
42. M. Anshari, M. N. Almunawar, S. A. Lim, A. Al-Mudimigh, Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*(94), 101 (2019)
43. G. Fotaki, M. Spruit, S. Brinkkemper, D. Meijer, Exploring big data opportunities for online customer segmentation. *International Journal of Business Intelligence Research (IJbir)* **5**(3), 58-75 (2014)
44. A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge* **12**(2), 1-257 (2021)
45. T. Le Dinh, T. T. Pham Thi, C. Pham-Nguyen, L. N. H. Nam, A knowledge-based model for context-aware smart service systems. *Journal of Information and Telecommunication*, 1-22 (2021)
46. Y.-H. Chen, C.-T. Su, A Kano-CKM model for customer knowledge discovery. *Total Quality Management & Business Excellence* **17**(5), 589-608 (2006)
47. N. A. K. Dam, T. Le Dinh, W. Menvielle, Key Aspects of Customer Intelligence in the Era of Massive Data. (Springer International Publishing, 2021)
48. M. J. Shaw, C. Subramaniam, G. W. Tan, M. E. Welge, Knowledge management and data mining for marketing. *Decision support systems* **31**(1), 127-137 (2001)
49. S. K. Taghizadeh, S. A. Rahman, M. M. Hossain, Knowledge from customer, for customer or about customer: which triggers innovation capability the most? *Journal of Knowledge Management* (2018)
50. J. Lu, L. Cairns, L. Smith, Data science in the business environment: customer analytics case studies in SMEs. *Journal of Modelling in Management* (2020)
51. N. Taherparvar, R. Esmailpour, M. Dostar, Customer knowledge management, innovation capability and business performance: a case study of the banking industry. *Journal of knowledge management* (2014)
52. M. K. Daradkeh, Determinants of visual analytics adoption in organizations: Knowledge discovery through content analysis of online evaluation reviews. *Information Technology & People* (2019)

53. A. Asllani, F. Luthans, What knowledge managers really do: an empirical and comparative analysis. *Journal of knowledge management* (2003)
54. A.-S. Oertzen, G. Odekerken-Schröder, S. A. Brax, B. Mager, Co-creating services—conceptual clarification, forms and outcomes. *Journal of Service Management* **29**(4), 641-679 (2018)
55. J. Trischler, S. J. Pervan, D. R. Scott, Exploring the “black box” of customer co-creation processes. *Journal of Services Marketing* **31**(3), 265-280 (2017)
56. N. A. K. Dam, T. Le Dinh, W. Menvielle, Customer Co-creation through the Lens of Service-dominant Logic: A Literature Review. *AMCIS 2020 Proceedings*. 29., (2020)
57. J. Forcier, P. Bissex, W. J. Chun, Python web development with Django. Addison-Wesley Professional, (2008)
58. M. Ahmed, M. M. Uddin, M. S. Azad, S. Haseeb, MySQL performance analysis on a limited resource server: Fedora vs. Ubuntu Linux. *Proceedings of the 2010 Spring Simulation Multiconference*, (2010)
59. E. Bressert, SciPy and NumPy: an overview for developers. (2012)
60. W. McKinney, pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing* **14**(9), 1-9 (2011)
61. J. Hunt: PyMySQL module. In: *Advanced Guide to Python 3 Programming*. pp. 291-301. Springer, (2019)
62. G. Hackeling, *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd, (2017)
63. A. Berti, S. J. van Zelst, W. van der Aalst, Process mining for python (PM4Py): bridging the gap between process- and data science. *arXiv preprint arXiv:1905.06169* (2019)
64. S. L. France, S. Ghose, Marketing analytics: Methods, practice, implementation, and links to other fields. *Expert Systems with Applications* **119**, 456-475 (2018)
65. S. Fan, R. Y. K. Lau, J. L. Zhao, Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research* **2**(1), 28-32 (2015)
66. A. Amado, P. Cortez, P. Rita, S. Moro, Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics* **24**(1), 1-7 (2018)
67. W. M. van der Aalst, Process discovery: Capturing the invisible. *IEEE Computational Intelligence Magazine* **5**(1), 28-41 (2010)
68. J. Rowley, The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information Science* **33**(2), 163-180 (2007)
69. M. Alavi, D. E. Leidner, Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 107-136 (2001)
70. P. Vassiliadis, A. Simitsis, S. Skiadopoulos, Conceptual modeling for ETL processes. *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, (2002)
71. U. Shafique, H. Qaiser, A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research* **12**(1), 217-222 (2014)
72. M. Drăgoicea, L. Walletzky, L. Carrubbo, N. G. Badr, A. M. Toli, F. Romanovská, M. Ge, Service Design for Resilience: A Multi-Contextual Modeling Perspective. *IEEE Access* **8**, 185526-185543 (2020)