O TRANSPONDER, WHERE ART THOU?

A Case Example of RFID Data Analytics in Retail

Jasser Al-Kassab
University of St. Gallen, Institute of Technology Management, St. Gallen, Switzerland
jasser.al-kassab@unisg.ch

Frédéric Thiesse
Chair of Information Systems Engineering
University of Wuerzburg
Wuerzburg, Germany
frederic.thiesse@uni-wuerzburg.de

Elgar Fleisch
University of St. Gallen
Institute of Technology Management, ETH Zurich, D-MTEC, St. Gallen/Zurich, Switzerland
elgar.fleisch@unisg.ch

Abstract—This contribution is concerned with the business value of large amounts of data generated by RFID data collection infrastructures. We present the case example of a department store that implemented RFID in its menswear department to seamlessly track thousands of items on their way from the distribution centre to the point of sale. The recorded trace data allow for a number of quantitative analyses, which provide novel insights into in-store logistics and customer behaviour on the sales floor. We provide a detailed overview of the underlying raw data, the deployed mechanisms for data filtering and aggregation, and the analysis procedures for generating business-relevant information. Our case study findings indicate that RFID poses an untapped opportunity for retail companies to improve category management, store layout design, inventory control, and process execution.

Keywords-RFID, retail industry, in-store logistics, customer behaviour.

I. INTRODUCTION

The growing interest in the use and application of Radio Frequency Identification (RFID) on the part of the retail industry in recent years has sparked an intensive debate in academia and practice on the benefits to be expected [1][2][3]. A large number of white papers, articles in trade journals, and research contributions have discussed the impact of RFID on supply chain performance. The benefits of RFID in supply chain operations described therein can be broadly categorized into (a) labour and time savings due to process acceleration (e.g., quicker pallet identification upon loading into a truck), and (b) benefits from increased visibility that lead to higher information quality or even process transformation [4][5]. The rationale of most existing RFID implementations follows the logic of the first category [6].

Against this background, this contribution is concerned with the second category of RFID benefits, i.e., the value of data generated by an RFID data collection infrastructure. Our research is motivated by the question, to what extent retail companies can draw benefits beyond simple efficiency gains from the analysis of large amounts of RFID data. For this purpose, we present selected results and lessons learned from a recent RFID project conducted by one of the largest department store chains in Europe. During a 15 month trial period, the company tested an RFID infrastructure in the menswear department of one of its stores with more than 30,000 tagged items on the sales floor. Their objective was to develop a better understanding of the manifold insights that could be provided by RFID into in-store logistics and customer behaviour. In the following, we present this example of an early adopter to shed light on the challenges and opportunities associated with the creation of business relevant information from RFID data.

II. RELATED WORKS

The various issues surrounding the processing of RFID data have been the subject of a steadily growing body of academic literature. Prior works can roughly be classified into four categories. First, several authors have discussed requirements and design alternatives for the implementation of specialized RFID middleware components to handle large amounts of raw data collected from distributed RFID readers. A second area of interest has been the design of algorithms for the filtering and aggregation of RFID data streams in order to derive interpretable information, e.g., on business events associated with RFID equipped products in the supply chain. Third, various researchers have proposed approaches for the efficient storage of RFID data, query languages and data structures, and other concepts related to data storage and management. A fourth research stream deals with the business value of RFID data in various industrial settings.

Our research contributes to the fourth category by a case-based study from the retail industry. Within this category, we are aware of only three works that discuss applications of RFID data analyses against the background of real-world implementations. This indicates that a gap exists in the literature regarding the practical use of RFID data and their business value. Delen et al. [7] identify a number of performance metrics that can be computed from RFID data and discuss how these measures can assist in improving logistical performance at a micro supply chain level of operations between a distribution centre and a store. Amini et al. [8] present a simulation study conducted at a hospital for which data related to trauma patient movement was collected with an RFID-based system. Baars and Sun [9] discuss options for modelling and utilizing multidimensional data sets for analytical applications using two case studies from retail and automotive. Our research differs from these prior works not only with regard to the sheer size of the RFID implementation.
under study but also in terms of the application domain and the level of detail of our analysis.

III. Case Background

We consider a European retailer, which offers a wide assortment of products with apparel and footwear in more than 100 department stores. The apparel retail industry in general is characterized by the management of a large variety of products, short life-cycles, high seasonality, high volatility, high-impulse purchasing, and complex distribution and logistics operations [10][11]. In addition, the company has particularly been facing strong competition in recent years by fast fashion retailers such as H&M and Zara, which continually increase their market share and yield significantly higher margins due to tightly integrated supply chains [12][13]. Customers have not only become more demanding and price conscious in parallel, their behaviour has also become increasingly difficult to assess, e.g., due to changing life patterns and the ageing of populations in Western countries [14][15].

In order to evaluate the technology’s potential to address the previously mentioned issues, the retailer implemented an RFID infrastructure in the menswear department in one of its department stores with a total of 22,000 square feet in size. Fig. 1 gives a schematic overview of the RFID installation. All products to be sold in the store are shipped from a nearby distribution centre (DC). Items are manually tagged in the DC and transported in trucks to the store, where they are read at the incoming gates. Goods intended for immediate sale are sent via the freight elevators to the menswear department, where they leave the back store and enter the front store via a transition gate.

On the average, 30,000 individual items equipped with transponder labels are constantly available to customers. Provided an item is not returned, the checkout reader usually poses the last part of the item’s lifecycle in the store. Besides customer returns, items can also flow back from the front store to the back store if they are stored in the local storage room, returned to the supplier or sent to other subsidiaries. RFID readers are installed at the escalators and elevators, on the gateways between the sales floor and the backroom, on several shelves, and in all 20 fitting rooms. In total, the infrastructure includes more than 60 RFID readers and more than 200 antennae. Some RFID readers are equipped with photoelectric barriers and motion detectors in order to determine the direction of the flow of goods, e.g., upon transit from the back store.

IV. Data Filtering and Enrichment

The continuous collection of data by the RFID infrastructure allows the retailer to conduct a number of analyses that go beyond what is already being done with the help of sales and inventory data. For the technology to be usable for this purpose, however, the raw data generated by reader devices have to undergo a number of processing steps. In order to derive meaningful information from large amounts of RFID data, the following three data sources were used:

- EPC event data collected by RFID readers are stored in a central ‘EPC Information Service’ (EPCIS) compatible data repository based on the family of ‘Electronic Product Code’ (EPC) standards [16]. EPCIS events are the data constructs stored in the EPCIS repository, which represent object detections over time [17]. All event types contain mandatory information on the date and time of an event, which objects or entities were subject to the event, where the event occurred, and in which business context the event happened.

- Master data are usually provided by the supplier or the retailer’s procurement department. They describe a product on a category level, i.e., each data set refers to one specific EAN code. This is in contrast to EPCIS event data, which refers to individual items. While the

Figure 1. Schematic overview of the RFID installation
structure of master data records is the same for all product types, the range of values for single attributes (e.g., size and colour codes) is supplier-specific, which complicates comparisons between products and categories.

- POS data are captured using barcodes, which encode traditional EAN numbers. Other data attributes are entered manually by sales staff. Despite the existence of master data, size information is collected redundantly since some products carry the same EAN code for all sizes.

Based on these data sources, a number of higher-order information sets were generated. On the one hand, filtering mechanisms were used in order to derive information on business events in the store from raw RFID data. The elimination of so-called ‘false positive reads’ (e.g., involuntary detection of items passing by) provides a typical example of a filtering mechanism. On the other hand, these events were aggregated and enriched by master and POS data, which provide the corresponding business context. The resulting information sets then allow for calculating various performance metrics (cf. Fig. 2).

In more detail, the combination of the three data sources was used by the retailer to generate the following information sets:

- **Trace histories** allow for reconstructing a product’s lifecycle over time, starting from the labelling point in the DC and ending at the check-out on the sales floor. In between, the product might be placed on a smart shelf, brought back into the back store, stored, returned to the smart shelf, and finally sold. Trace histories of this kind are necessary for all analyses in the area of process execution (e.g., process cycles and promotion execution), category management (e.g., lead times), and inventory management (e.g., out-of-shelf situations, out-of-store situations, and overstocking). In order to generate trace histories, each item’s read events are grouped and sorted by the corresponding time stamps. In a second step, the read events of the product’s movement between back and front store are transformed into a linked list by interconnecting the individual read events using the attributes ‘ReadPoint’ and ‘eventTime’, using knowledge of the store layout and the RFID installation. In a third step, the temporal delays between individual read events are calculated by subtracting the current event time from the previous one. Since the front store and back store location information of each read point is known, lead times, inventory, and process execution analyses can be conducted based on the resulting table.

- **Smart shelf inventories** are generated from RFID-enabled shelves (so-called “smart shelves”) on the sales floor, which constantly interrogate tags and update their inventory level in the EPCIS repository. In order to limit the amount of data to be stored in the database, the periodicity of EPCIS updates was set, depending on the smart shelf, to 15 or 120 minutes, respectively. From these smart shelf read events, inventory data are generated that allow for tracking inventory changes over time. For this purpose, a table only containing read events from smart shelves is created. Besides analyses of ‘in stock but not on shelf’ situations, these data allow the identification of misplaced merchandise, since they read all items placed on the shelf, irrespectively of whether they are supposed to be on that shelf or not.

- **Try-ons** are investigated in order to monitor the utilization of the 20 fitting rooms on the sales floor, which are grouped into 5 clusters. For this purpose, the fitting room reads are extracted from the EPCIS. In a second step, four different types of reading errors have to be removed from the data: (i) fitting rooms reading items carried by passing customers or employees, (ii) items read by adjacent cabins, (iii) left behind items, and (iv) items from nearby merchandise fixtures. From a filtering perspective, case (i) is comparable to case (ii), as case (iii) is to case (iv). For the cases (i) & (ii), an algorithm was designed that filters these reading errors by deleting the events of a certain item that

![Figure 2. Overview of data analysis procedures](image-url)
occurred within a certain time span in other (adjacent) fitting rooms. As for cases (iii) and (iv), the events of the corresponding items are counted only once, even if they were read for hours. The resulting table comprises a complete list of individual try-ons including the products that were taken to the fitting rooms.

- **Merchandise performance** denotes the correlation between sales, try-ons, and inventory levels on the sales floor. The try-on events described in the previous subsection are combined with POS data and the inventory level at the time of the try-on. A challenge was the attribution of fitting room visits to a purchase. Since customers could not be identified by RFID, a probabilistic algorithm was designed that would group try-on events that occurred within a certain time span and assign them to one fictitious customer. For this purpose, the fitting room visits of each fitting room are selected and ordered by time. Then the algorithm would iterate through the list and group the events by time for each customer.

V. **PERFORMANCE METRICS**

In this section, we illustrate how the retailer under study made use of the previously gathered RFID data to calculate several performance metrics based on raw, filtered, and enriched RFID data to gain insights into his internal processes and customer perceptions of his store assortment.

A. **RFID Reader Performance**

Analyses of RFID raw data alone do not provide much business value but help the company to monitor the performance of their data collection infrastructure and the quality of the generated data. By comparing POS data with RFID data from the check-out reader, for instance, it is possible to investigate the difference between the two data collection procedures, which allows for drawing conclusions on the fraction of undetected products (cf. Fig. 3) and hence shows room for improvement. Other examples are periodic tests of the performance of their data collection infrastructure and the business value but help the company to monitor the quality of the generated data. By comparing POS data with

B. **Front store / back store inventory**

Using trace histories, the front store and back store inventory of items can be calculated in real-time for a given time and date. For example, a list can be generated for items on EAN level, for each day of the year and plotted as a graph (cf. Fig. 4). By providing more visibility regarding the distribution of the department’s inventory stock between front store and back store, which was not possible in that way before the introduction of the transition gate readers between back and front store, this analysis supports the retailer in reducing the number of out-of-shelf but in stock or stock-out situations, which are responsible for lost sales and reduced customer satisfaction. Moreover, this analysis can help to detect promotion execution errors (i.e., promotional articles, which are put on the sales floor too early), and storage inefficiencies (i.e., low inventory levels on the front store, but high inventory levels in the back store).

![Figure 4. Front store and back store distribution of articles over time](image)

C. **Process Cycles**

In regards to process execution, the retail company wanted to assess the efficiency of their in-store processes. They therefore analyzed the event data and looked for certain patterns, such as loops, or products that were taken very often to the fitting room, or which moved between different floors. For that purpose, trace histories are aggregated in order to count the number of an item’s appearances at different read points, such as fitting room readers, readers between front store and back store, elevator and escalator exits, storage, and checkouts (cf. Fig. 5).

![Figure 5. Process cycles (different items/EPCs in comparison)](image)

This analysis provides the retailer with a higher process visibility and enables him to detect previously unobserved inefficiencies. These product pedigree analyses require individually tagged items, be it RFID or the barcode. However, RFID allows for a higher degree of automation as compared to the barcode which requires a direct line of sight for item readings.

D. **Misplaced merchandise**

The issue of misplaced merchandise situations is mainly caused by customers not returning the items they tried on to the original merchandise fixture. They would either leave the items in the fitting rooms or on nearby shelves. In order to detect misplacements, inventory data from the smart shelves are searched for items that were not supposed to be on the shelf. In a second step, these misplaced items were grouped according
to their affiliation with their department and the duration of the apparent misplacements (cf. Fig. 6).

![Figure 6](image_url) Occurrence of misplaced merchandise on the smart shelf by category (above); duration of misplacement situations (below)

All items with short misplacement duration stem most probably from customers passing by the shelf (false positives); items with very long misplacement duration are most probably constant and remote readings from nearby fixtures. From the remaining item data in-between, the retailer gains a picture of the cleaning and tidying processes on the sales floor. Studies exist that examine the degree of product misplacement in retail stores [21]. The data, however, is collected manually in that case. Automatically generated and gathered RFID data allows for more fine granular analyses.

E. Stock-outs on the smart shelf

Besides the detection of misplacements, inventory level data from smart shelf readings also allows for the detection of stock-outs. Technically, this analysis is equivalent to the back store / front store analysis previously described. However, the underlying data quality is higher since RFID transponders on the respective items are interrogated permanently, i.e., the probability of reading errors is significantly smaller. Furthermore, the level of detail is higher since stock-outs cannot only be detected on the level of sales floor as a whole, but rather be attributed to an individual shelf.

F. Utilization of fitting rooms

In order to analyze the utilization of fitting rooms, try-on data are counted and grouped into clusters (cf. Fig. 7). With the information provided by this analysis, the retailer under study can reassess the trade-off between the total number of fitting rooms on the shop floor and within one cluster – with a surface of 16 square foot each – against more sales floor space. Besides the overall utilization of the clusters, the analysis could also point out the utilization during peak times, thus helping the retailer to either eliminate some of the fitting rooms in the clusters in favour of sales floor space, or in favour of fewer, but bigger fitting rooms respectively. Again, this analysis is only economically feasible with automatically collected RFID try-on data.

![Figure 7](image_url) Utilization of the fitting rooms (FR) and clusters

G. Lead Times

Trace histories can also be aggregated on the EAN level and on the category level (cf. Fig. 8).

They can be used to investigate lead times with regard to different lifecycle stages of a given product. Lead time analyses support the retailer in improving his category management and his process efficiency. Articles that spend larger periods of time on the sales floor until they are sold, for example, could be removed from the store’s assortment. The same applies for unsellable sizes, colours, or brands. When these analyses are carried out on the item-level, all products that spend more time than the average in storage or on the sales floor, give indications that they are either lost in the storage, stolen, or that they are damaged or broken, and can thus be physically removed from the sales floor or virtually deleted from the merchandise management system. The lead times can be shown for all areas of the supply chain. They can therefore also be used to optimize supply chain efficiency.

H. Catchment area of fitting rooms

Fig. 9 depicts the layout of the menswear department, subdivided into shop-in-shops for different article groups and...
brands. Just like each merchandise item on the sales floor, each merchandise fixture and each shop has a unique identifier. Each item is assigned to a merchandise fixture, and each fixture is assigned to its shop. This additional information can be used to visualize the catchment areas of fitting room clusters. The graphical visualization of catchment areas on the sales floor may help the retailer to optimize the sales floor layout, the fitting room positioning, and give indications about movement patterns of customers. This information can also be used to optimize product placements on the sales floor, especially of complementary articles, in order to influence customer shopping paths. This new level of transparency is now newly possible with the try-on read events in the fitting rooms.

Figure 9. Comparison of the catchment areas of three different fitting room clusters

J. Correlation between sales, try-on frequency, and inventory

During the data filtering and enrichment process, sales information, try-on information, and inventory level information for each article and read event are merged into one data table. As such, these data can be analyzed using logistic regression models in order to identify a level of inventory on the front store that would be optimal in terms of number of try-ons and sales, and eventually customer satisfaction. With these results, the retailer under study can systematically approach the optimal level of articles on the front store, thus increasing sales and customer satisfaction, and at the same time decreasing the capital costs of items on the sales floor. For the optimization of his category management, the number of try-ons and sales for each article group are correlated (cf. Fig. 11), resulting in a try-ons/sales ratio. With this information, the retailer can compare articles within one article group by brand, supplier, colour, and size, for example, thus optimizing his product range by removing articles from the assortment that were often tried on but rarely bought. The assortment can then be reassessed and improved. Items that are sold with no or little try-ons can also be shelved close to the checkouts, for example.

Figure 11. Correlation between try-ons and sales

K. Staff allocation

Based on the same data, the retailer can, for example, compare try-ons over time with the subsequent sales events in order to optimise the number of employees on the sales floor. The chart depicted in Fig. 12 shows the number of try-ons for a given article group during the opening hours of one day, the number of sales events, and the ratio of both values (represented as bars). The data indicate that the ratio of try-ons and sales events changes significantly over time. In this case, the retailer under study could not profit proportionally from the rush of interested visitors, trying on articles in the early evening times.

Figure 12. Try-ons / sales ratio over time (i.e., opening hours)
The comparison of the try-on/sales ratio at 3 pm with that at 5 pm shows that an increase in try-ons by more than 100% could not be translated into a corresponding sales increase. This might be a consequence of insufficient staff on the sales floor during these hours, resulting in lost sales, especially for articles that require extensive customer counselling, such as suits, for instance.

VI. CONCLUSIONS

This contribution provided a case example of an innovative and innovation-driven RFID project at one of the largest department store chains in Europe. The objective of our study was to investigate the opportunities opened up by RFID data in retail beyond efficiency gains by means of simple process acceleration. The case directly points to one of the fundamental and largely unresolved questions in the IS literature: how can organisations generate value from information [18]? With the help of several illustrative examples we have shown how a company can combine RFID with traditional data sources in order to generate many novel performance metrics that create unprecedented insights into in-store process execution and customer behaviour on the sales floor.

Our case example supports the hypothesis that RFID holds the potential to create several informational impacts on the management processes of a department store chain (cf. Table 1). On the one hand, RFID provides detailed information on the reality of store process execution, which can be used for continuous process improvement. Detailed tracking of product movements between the DC, the back store and the front store, for example, help the retailer to eliminate previously undetected inefficiencies in his internal operations. On the other hand, the data collection infrastructure serves as a tool for better understanding aspects of customer behaviour and perceptions of changes in a store’s assortment, product presentation, or floor layout. Thus, RFID becomes a valuable data source for a number of management processes, e.g., category management, store layout design, inventory control, and process execution. Fig. 13 provides an illustration of the integration of RFID in combination with POS and master data into the retailer’s management cycle [19]. On the operational level, store and customer processes are monitored by collecting the according data using the RFID infrastructure and the existing barcode-based checkout system. Based on these data together with master data extracted from the inventory management system, analyses as described in this paper can be conducted. The results allow for more insights into operational processes and thus to put in place an efficient and effective process control (see also [22] and [23] for a categorization of effects of RFID technology on the operational and managerial process level). For instance, the “stock-outs on the smart shelf” analysis proactively and automatically informs employees on the sales floor about out of shelf situations. Sales employees no longer need to constantly check the inventory levels on the shelves in a paper-based process; corresponding replenishment activities are automatically triggered. Similarly, the “misplaced merchandise” analysis indicates that a shelf has to be tidied up. On an aggregated level, the results of the data analyses improve the managerial processes in so far, as they lead to incremental process change and process innovation. For instance, the “complements and substitutes” analysis, which identifies which articles are often tried on together in the fitting room, will lead to an improved store layout design, since these articles will be positioned close to each other on the sales floor. Or, the “staff allocation” analysis will lead to improvements in the staff allocation on the sales floor, since the retailer knows about the customers’ footfall on each day of the week and on each hour of the day.

While the value of these data seems evident and compelling at first, the construction of a cause-and-effect chain between RFID investments and an increase in process performance is not a trivial task [20]. In our case example, the value of continuously analyzing data collected by the RFID infrastructure was affirmed by virtually all involved project team members. However, a general answer on how to make use of this information to generate quantifiable impacts on performance could not be given so far. This finding leads us to the conjecture that RFID impacts on the level of management processes depend on the existence of complementary and contextual factors (e.g., information capabilities) that are required to translate RFID data into value. Against this background, we see opportunities for further research in the area of RFID.
various directions. Firstly, our study is limited by the fact that it draws upon only one single, not necessarily typical case example of an early adopter. Additional cases of real-world implementations are indispensable to further examine the value of RFID data to business processes.

On the long run, longitudinal studies will be necessary to confirm our assumption that RFID data can positively influence managerial decision making and, thus, lead to some form of competitive advantage. Secondly, more research should be directed to the development of sophisticated models and tools that support the decision maker in the design of RFID systems and processes. Moreover, the current gap in the body of literature regarding complementary and contextual factors of RFID implementations should be filled by empirical works that analyze the nature and the role of a firm’s capabilities and performance.

On the long run, longitudinal studies will be necessary to confirm our assumption that RFID data can positively influence managerial decision making and, thus, lead to some form of competitive advantage. Secondly, more research should be directed to the development of sophisticated models and tools that support the decision maker in the design of RFID systems and processes. Moreover, the current gap in the body of literature regarding complementary and contextual factors of RFID implementations should be filled by empirical works that analyze the nature and the role of a firm’s capabilities and performance.

**ACKNOWLEDGMENT**

The authors want to thank Mr Marco Hornung for his contributions in documenting the data set and in overcoming the practical difficulties in its treatment.

**REFERENCES**


