Multidimensional Analysis of RFID Data in Logistics

Henning Baars
Chair of Information Systems I,
University of Stuttgart, Germany
baars@wi.uni-stuttgart.de

Xuanpu Sun
Chair of Information Systems I,
University of Stuttgart, Germany
sun@wi.uni-stuttgart.de

Abstract
In the domain of logistics, Radio Frequency Identification (RFID) promises a plethora of benefits due to an enhanced efficiency, accuracy, and preciseness of object identification. Many of the far-reaching solutions proposed in the literature are based on data exchange, integration, and analysis. However, the respective applications have so far not been thoroughly scrutinized regarding relevance and design. This paper specifically explores options for modeling and utilizing multidimensional data sets for analytical applications as known from the realm of “Business Intelligence” (BI). Based on case studies, analysis scenarios are derived and discussed with respect to characteristic requirements for the purposeful utilization of BI tools. This is the foundation for the design of a multidimensional data model and for the construction of a prototype based on Online Analytical Processing (OLAP). The prototype not only visualizes possible applications and their business potential, its construction also leads to findings regarding the gathering and integration of RFID data.

1. Motivation and relevance

At its core, Radio Frequency Identification (RFID) automates object identification. It allows for a contact-free, radio based data capturing either for logistical units (pallets, boxes etc.) or for individual items [1, 2, 31]. Relevant business potential stems from the possibility of bulk identification, options to store data along with the ID, and the combination of RFID with sensor technology and/or with tracking and tracing systems [10, 11].

In the domain of logistics, the possibility to efficiently gather object data with unprecedented accuracy and frequency in real time opens the door for a number of applications that aim beyond mere automation scenarios [25, 28, 36]: By increasing the number of measuring points and measuring events, the overall supply chain visibility is increased. This enables better decisions (“informational benefits”) and even far-reaching process redesigns (“transformational benefits”) [6, 9, 33]. The business relevance of this type of applications has not only been claimed in various publications [32, 36], but also been shown in first empirical studies among pilot users [20].

The accompanying need for an exchange of RFID data across enterprise borders is addressed by the implementation of infrastructures like the EPCglobal network [17, 19]. It has already been illustrated that there is a distinct need for further data integration, harmonization, and aggregation in such environments [35]: Many of the discussed monitoring or analysis scenarios require binding together data from a variety of heterogeneous operational systems owned by different organizational units or even different enterprises. Moreover, to identify patterns, data needs to be aggregated in numerous ways.

The resulting infrastructural consequences for data refinement and analysis have not yet been scrutinized: It remains open what concrete types of systems enable meaningful insights into RFID data, what requirements regarding data preparation they bring along, and which data models they can be based on.

This contribution specifically explores how enriched RFID data can be prepared in an integrated, multidimensional fashion for purposes of managerial support. *Multidimensional data sets* are specifically designed for the analysis of large data volumes and for flexible aggregations along multiple dimensions (e.g. product, time, location, etc.). Those dimensions define the context of the respective indicators, called “facts”. The potential of this approach for RFID data is demonstrated by the use of an Online Analytical Processing (OLAP) tool. OLAP denotes a data analysis approach that is based on the navigation in multidimensional data sets. It allows the user to freely select facts, dimensions, and granularity layers in an ad-hoc fashion and to immediately visualize the results [12, 23, 27].
Requirements for a management oriented analytical support are drawn from two case studies. This leads to the derivation of an integrated, relational, multidimensional data model (star scheme) and – based on that – the design and the development of an OLAP-based prototype. The prototype serves for illustration purposes and supports a further investigation of the possibilities of a multidimensional analysis of RFID data. The data modeling and prototype building activities also generate insights into the requirements for an underlying data transformation and integration architecture.

2. Related work

In recent years, RFID has already drawn much research attention. A notable result from case and pilot studies is that – despite some challenges – RFID implementations in the logistics context are indeed feasible and that some of them bring along real world business potential [18, 24, 32].

Various publications emphasize the relevance of informational and transformational benefits: Results underpinning this assertion come from simulations, case study research, and empirical studies among pilot users [20, 25, 28, 29, 36]. This class of effects gains particular in weight when it comes to large-scale RFID implementations that involve multiple sites or even several independent organizations. The rationale behind this becomes clear when put into the context of the general theme of data sharing along the supply chain – the relevance of which has been shown numerously before [3, 8, 22].

The subset of applications that is discussed here is based on refined data, which means that data from various sources is harmonized, aggregated, and enriched according to managerial needs. A potential starting point to tackle the respective tasks is to apply established solutions from the domain of Business Intelligence (BI) – which denotes approaches for integrated managerial support [4] – to the RFID and logistics area, namely solutions built upon Data Warehouses.

A Data Warehouse (DWH) is understood to be an integrated, non-volatile, subject-oriented data collection for decision support [23]. Based on the DWH, systems for data preparation and analysis can be installed: Among others, various sorts of reporting solutions, data mining systems for statistical pattern recognition, ad-hoc-query tools, and OLAP applications for multidimensional data analysis [4, 23, 34].

DWH-infrastructures are also increasingly used for the support of operational monitoring and near time analysis scenarios. The respective real time and active DHWs are turning into logically centralized data refinement hubs that support all managerial levels and business functions.

The combination of BI and RFID has already been accessed from a technical angle [21, 30, 35]. The focus of the respective research activities especially lies on the consequences of storing RFID data on transactional granularity in data warehouses and approaches for handling the resulting data volume.

There is still a lack of insight into an application-oriented preparation of RFID data. This paper concentrates on OLAP systems because they vividly illustrate the potential of keeping a rich multidimensional data collection in the DWH that is designed for flexible uses.

3. Methodology

The requirements for the data modeling and the prototype design are derived from two case studies. The first one comes from the realm of retail logistics while the other addresses production logistics in the vehicle manufacturing industry. Both case studies applied a variety of qualitative data gathering techniques (narrative interviews, workshops, observation, analysis of documents) which were used to gain insights into the benefits of possible combinations of BI and RFID, to derive and evaluate relevant analysis scenarios, and to elicit respective data provision and integration needs. In both studies the involved experts and managers already built up on pilot-study experiences with RFID technology.

The results of the case studies are analyzed for relevant characteristics with respect to a possible system design. This leads to the formulation of seven requirements which are guiding the design of the data model and the prototype. The fit of the outcomes with the requirements is demonstrated with selected analyses from the supply chain case. The construction of the data model and the prototype also entails some relevant consequences for data refinement.

4. Deriving scenarios and requirements

The following two case studies are the foundation of the subsequent requirements definition:
4.1. Retail chain case

The first case was accomplished within a logistics project of an European retailing group. It addresses coupling EPCglobal infrastructures with systems for management support, as this was seen as highly relevant within the project consortium. The derived scenarios reflect the actual business demands, envisioned applications, required data, and possible benefits of concrete analyses [5].

Among others, the chain in discussion includes Chinese manufacturers, a Goods Distribution Centre (GDC) in Germany (handles the goods transfer to the individual retailers), as well as retail outlets across Europe. Assumptions being made for the derived scenarios include the implementation of the EPCglobal infrastructure alongside the introduction of a "Vendor Managed Inventory" approach (cf. [14]), and a far reaching redistribution of decision rights to the GDC. The GDC is turned into the core decision unit in the chain for distribution related activities. For some of the discussed analyses, a successful introduction of RFID tags on item level and/or the coupled use of RFID tags and sensors are assumed.

Predicated on these presumptions, data collection points were derived and discussed for each partner in the chain. Furthermore, the availability of additional data sources was elicited and scenarios for RFID data usage including their respective data provision requirements were developed.

From the perspective of this paper the following analysis scenarios are deemed to be relevant:

- Maximization of shelf space utilization
  In this first scenario, the actual amounts of goods in the shelves are taken into consideration within disposition decisions of the GDC for purposes of a maximization of shelf space utilization. The effectiveness of the resulting replenishing policies can be tracked with current data and analyzed with historic data.

- Relocation of goods between retailers
  The GDC becomes responsible for balancing out shortages by directing goods transfers among different retailers.

- Identification of root causes for defective articles
  This scenario aims at identifying root causes for quality issues. This requires RFID tags on item level, gathering tracking and tracing data, as well as the inclusion of sensor information.

- Demand driven order process
  The aim is to leverage a more demand driven order process to realize reduced lot sizes and a higher order frequency – coordinated by the GDC and based both on real time and historical RFID data.

- Alternative transport routes
  The last analysis pertains to route planning. In the given case, RFID, GPS, and sensor data enable the GDC to monitor the status and the location of moving goods and to be able to intervene as soon as possible. The selection of alternative routes can be based on collections of historical data.

Note that although most of those problems are known from supply chain management literature (e.g. [15, 16]), they could only partially be tackled in the given environment due to issues of data availability (for example sensor data, item based information, detailed in-house localization), data quality (precision, correctness, timeliness), and especially data integration – all analyses require to bind together data from at least two involved parties [22, 37].

4.2. Vehicle production case

In the second case, internal logistics processes within a vehicle manufacturing setting are focused. The scenarios in discussion are addressing the production and transportation of a core vehicle component and the management of loading units. As in the first case, both relevant RFID measurement points and scenarios for the analysis have been derived together with involved experts from the company. It needs to be stressed that the discussed issues are widespread among similar sized manufacturing companies [7, 13]. The study provided concrete insights into the qualitative nature as well as in the relevance and the feasibility of the conceived solutions [26].

From the viewpoint of this paper the following analysis scenarios are noteworthy:

- Rejects over time
  This analysis binds data from different systems together in order to track and analyze rejects over time for the adjustment of production planning and materials provision.

- Analysis of throughput times
  To support the optimization of production rates and logistics planning, throughput times are analyzed for the complete processes as well as for dedicated sub-processes. Among others, the involved machines can form a relevant analysis dimension.

- Alternative transportation routes
  Similar to the retail case, the transportation to different (production) sites needs to be monitored regarding quality and time – if critical deviations occur, alternative routes and/or transpor-
4.3. Deriving requirements for a Management support system

From the management support perspective, the following aspects are relevant:

- **Volume of loading units / shrinkage**
  For the optimization of the volume of loading units, reliable information on the actual amount and distribution of available units is required. This includes the need to view selections of the total data set with respect to the unit types, locations, and responsible organizational units.
  The analysis also addresses pinpointing root causes for losses in the loading unit supply. As in the first case, all analyses draw data from different systems. Many are based on item visibility across the complete process and require an item-based tagging. Bar codes are not considered to be a viable option due to the harsh environmental conditions and object handling costs. In the case of the loading units, the currently available data is completely based on daily manual counts.

- **Identification and measurement events**
  The proposed system in this paper is constructed therefor (stock keeping).
  An EPC-Code can be understood as a natural primary key that is valid across enterprise borders. The integration potential of RFID is the basis for several analyses like the localization of root courses for defects, for the comparison of alternative routes backed by historical data, for pinpointing shrinkage origins, etc.

- **Data capturing**
  An efficient data capturing also allows for more frequent and more precise measurements – and therefore for a higher temporal and spatial resolution during an analysis: The data can be sliced with greater precision. An example is the differentiation between goods in the retail shelves and in the retail warehouse for maximized shelf space utilization.

- **Automatic measurement**
  The automatic measurement leads to a higher data quality and thereby a higher decision quality. This makes it possible to cut down security buffers kept for coping with uncertainties – both in internal and external logistics.

- **Feedback cycle with planning**
  Usually analyses are embedded in a mutual feedback cycle with planning – planned measurements are mirrored against current data and plans are adjusted according to actual (measured) events. Examples include the selection and adjustment of routes, or the optimization of throughput times and buffers.

- **The identification can be tied to additional measurements**
  The identification and measurement events need to be put into their logistical context – which at best means the association with the complete path from the upstream origin to the downstream destination. In consequence, the events need to be qualified with information on the route and their position in the route. In the logistics domain, the relevant events captured by RFID are obviously for the most part physical movements along an overarching global route or the lack thereof (stock keeping).

  **Requirement 1**: The system should allow analyzing transportation events within their respective logistics context.

  The spatial resolution that can be utilized is determined by the measurement points (RFID readers) – with the most granular level being the distance between two adjacent readers. In the discussed logistics context, the set of meaningful pairings of readers is determined by the sections of the global route. The respective events are transfers of items along those track sections.

  **Requirement 2**: The system should allow drilling down to individual transfer events which are delimited by a track section on a global route.

  Although the relevant time resolution can be made almost arbitrary small by increasing measurement frequency, meaningful insights are only delivered when either a transfer has been completed or an unexpected change of the status of a transported item or logistical unit takes place, e.g. when an item is lost or rejected due to a defect.

  **Requirement 3**: The system should track the current status of an item.

  For insights into the current distribution of goods and their status, it is necessary to summarize amounts
of goods with respect to their current position, while for the historical analyses the number of completed or failed transfers are the bases of the calculations.

**Requirement 4:** The system should be able to provide insights into current as well as historical data.

As discussed above, it is recommendable to always mirror actual events against planned ones. This includes the occurrence of item transfer events as well as the performance measures that go along with those events, like throughput times, costs, etc.

**Requirement 5:** For managerial purposes, deviations between actual and planned performance indicators as well as actual and planned events need to be tracked.

The last two requirements stem from the necessity to adequately qualify the data in order to draw valid conclusions from it:

**Requirement 6:** The system should provide an array of relevant performance indicators that qualify the transfer events.

**Requirement 7:** The system needs to integrate a large set of descriptive dimensions and attributes to describe the context of the indicators.

5. **Data model**

The data model is strictly based on a relational, multidimensional design (“star scheme”) [27]. It is portrayed in Figure 1. The model is deliberately denormalized for reasons of simplicity and compatibility (avoiding so-called “snow-flaking”).

The relation item transfer gathers the facts that need to be analyzed. Each transfer event refers to an individual track section – thereby fulfilling the first two requirements. The item transfer can on the one hand be conducted using different shipment modes and qualities and on the other hand include item modifications like manufacturing steps, or stock keeping. This differentiation is described in the mode of transformation dimension.

The track section relation also stores the logistic context – it assigns each section to a sub route and to a global route which also define the aggregation paths. The sub and global routes are only defined by their start and end points and can be used to track an arbitrary number of routing options. By coupling the item number and the track sections items or groups thereof can be followed throughout the complete supply chain, making way for analyses of the sort

![Figure 1: Structure of the data model](image-url)
“How many items reached their destination?” or “What number of items went which way on a global route?” The ordinal position on a route is denoted by a sequence number. This enables meaningful data selections.

The status demanded in requirement 3 is kept in a dedicated field of the fact table. The status can also capture the current state of an item that has not yet reached its destination. By using the status field to exclude arrivals, an aggregated near time monitoring is enabled (requirement 4).

Providing planned data to confirm with requirement 5 could not be realized in the classic way by simply including budget data in the fact table: Each item transfer is conceived to be an actual event – defined by a real item and a real track section at an actual period of time. However, the actual route a item takes can very well deviate from the plan. For this reason, the field planned flag is included: The planned data is kept in separate fact fields that can be mirrored with the actual data by a join over the unique item number.

Requirements 6 and 7 are saliently depicted in the model. Of particular interest are the box, pallet, and container dimensions which store packaging information: Each individual article is an item that can be assigned to such a logistical unit during an item transfer. Simultaneously, the logistical units themselves are treated as items, although they have different “global routes”.

This data model does not include any recursions for packaging and routes in order to impose uniform aggregation paths. This leads to the requirement that such structures from operational systems must be recursively taken apart during data extraction. Figure 2 shows the three level hierarchy that has been designed for the prototype regarding the route:

![Figure 2. Three level hierarchy](image)

The global route labels the complete actual or planned transfer from the initial origin to the ultimate destination, while the sub route denotes a part of that route. This middle layer is included to be able to group subsequent track sections based on organizational responsibility. The section is the smallest unit – the path between two RFID readers. For large scale implementations, the hierarchy might be further divided up. This should especially be necessary when a multitude of small-scale transport events enters the data set.

Eventually, quite a few attributes on the dimensions are basically potential cost (or time) drivers. Their combination should be reflected in the facts.

6. Prototype and demonstration scenarios

The case studies introduced in chapter 4 were the starting point for the development of a prototype that is essentially built on top of the discussed data model. The prototype is implemented in MS SQL Server 2005 (representing the DWH) and the OLAP tool Cognos 8 (for the analyses). The data for the prototype is generated with a data generator that has been realized in Java. During the data generation process, systematical deviations can be injected into the data set (e.g. a tendency to take alternate, expensive routes, shrinkage, or temperature-induced spoilage). All relevant parameters regarding the set up and the deviations can be set in a XML definition file.

The scenarios that will be presented are based on the retail chain case and are built upon the same assumptions. All analyses work with a single integrated (play) data pool. It is assumed that a Chinese manufacturer delivers products to a German retailer (Figure 3). The applied “modes of transformation” include transportation by trains, ships, airplanes (on an alternative route), and trucks.
The current data pool addresses five different products with 10,000 items and covers a time horizon of one year. All planned and actual transfer events for the defined global route were generated. The resulting data volume already exceeds 40 MB. This indicates the amount of data that needs to be analyzed in a real world setting with several ten thousand product variants. The data volume results mainly from the item based tracking and the level of detail that is generated with the track sections. It primarily resides in the fact table.

In the following presentation, the problem area of transport route selection has been chosen because it surfaced both in the manufacturing and in the retail case. Furthermore, it exemplifies the potential of putting data into its logistical context, monitor current and historic data, drill into the data, and contrast plan and actual values. The problem thereby contains the core requirements discussed in section 4.3.

### 6.1. Monitoring

The data model enables a near time monitoring of item subsets which have not yet reached their target destination. It is especially possible to watch the number of items which have been rejected (manually or automatically sorted out because of defects or surpassing critical sensor values) or declared lost (identified as missing, e.g. with smart shelves). The OLAP tool allows conducting aggregations for defined sets of items (e.g. selected by product type and for a given global route etc.).

Figure 4 shows the total number of rejected and lost items for the actual and planned sub routes on the global route Beijing→Stuttgart.

The OLAP-based navigation on top of the derived data model allows for an immediate drill down and slicing of the data to give an overview on the current situation **within** one selected retail store. Thereby a clearer picture of “hot spots” can be gained. Figure 6 shows such an analysis. The sequence of tracksections follows the basic store layout from the case study as shown in Figure 5.

In the shown scenario the rejections agglomerate in goods reception and stocking. It can now further be analyzed if the situation is confined to a subset of the products, supplier, transportation mode etc.
Based on the discussed data model, the system also allows for a visualization of current bottlenecks, congestions, route deviations, and many other relevant events.

It needs to be emphasized that this type of scenario requires an efficient, near time data capturing mechanism on a level of detail and preciseness which is far from being available with established identification technologies.

6.2. Analysis of historic data

The storage and aggregation of historic data is a facilitator for the identification of patterns e.g. regarding spoilage or unstable delivery reliability. Among other scenarios, it supports the evaluation of different options regarding transportation routes.

For the demonstration, an analysis based on contrasting the actual and the plan number of items reveals a systematic deviation from the planned route – about 50% of the articles are taking an unplanned transportation route which includes air-transportation. The results are also reflected in the associated values for cost and duration.

Figure 7 depicts a snapshot that shows the number of items on its route and the resulting costs – for the complete global route and broken down on sub route level. Especially the inclusion of sensor data allows to extend the evaluation to a variety of non monetary measures, e.g. on average transport conditions.

For well-structured problems the respective data can also be used as a flexible and precise socket for model-based systems that exploit the abundance of relevant optimization algorithms.

7. Conclusions and outlook

As shown, the discussed data model incorporates the requirements introduced in section 3. Next to the scenarios presented in this paper, several others have already been filled with demonstration data. Although concrete dimensions and performance indicators vary, the basic model fits both the manufacturing and the vehicle case.

By aiming at integration and by including aggregation hierarchies and anchors for a variety of slicing options, the model is conceived to be adaptable to all sorts of applications that aim at the monitoring and analysis of complex good flows with a multitude of individual steps and a lack transparency due to complexity, volume, and geographical as well as organizational distribution. The approach is especially
suited for automated data collection scenarios which deliver large amounts of precise data in near time.

This again highlights the role of the RFID technology and standards like those defined by EPC-global, which are also seen as facilitators for the data integration across enterprise borders.

The role OLAP plays in this research is twofold: On the one hand, OLAP represents an established and powerful analytical approach by itself. On the other hand, during the iterative design of the model and the prototype it has proven to be a viable option to visually gather insights into the potential of an integrated, multidimensional data collection.

This data collection is seen as a core building block of a DWH with RFID data for applications in logistics. The DWH approach allows to embed analytical RFID-applications into centralized and specialized BI units and thereby to align them with the general trend in the decision support domain towards increased professionalization. However, from the data model it can already be seen that the refinement steps for the raw RFID measurements feeding the DWH are by no means trivial.

They include:
1. Assigning RFID measurement events to units of analysis – in the model: the track section. This also entails the necessity for a purposeful grouping of RFID readers.
2. Deciding on what measurements are needed for which analysis at what time – with consequences regarding the transfer medium (satellite, cell phone network, or fixed line) and the middleware infrastructure.
3. Aligning the track sections with the planned route.
4. Binding together events measured with RFID and data captured by other means, e.g. when sorting out defective products.
5. Harmonizing the measurement values syntactically and semantically for all involved organizations.
6. Harmonizing the dimensional data to get meaningful results.
7. Building up the aggregation hierarchies which include packaging and routing hierarchies. It needs to be considered that the actual and planned routes can deviate and the routes and the aggregation hierarchies for the track section need to be (re)constructed automatically – after each data load.

Upcoming research activities need to address this ETL-process and provide for suitable ETL tool designs. Besides, the prototype, the data generator, and the resulting data set are subject to continuous expansion and refinement.

Parallel to this, further research activities aim at a quantitative exploration of the demand among different types of firms as well as at an extensive interview-based validation for both the prototype and, based on that, the scenarios it embodies.

During this process, the prototype not only serves as a model for a potential analysis system. It also becomes a communication and research tool that visualizes conceived RFID scenarios and makes them falsifiable.

8. References


