

Increasing Supply-Chain Visibility with Rule-Based RFID Data Analysis

RFID technology tracks the flow of physical items and goods in supply chains to help users detect inefficiencies, such as shipment delays, theft, or inventory problems. An inevitable consequence, however, is that it generates huge numbers of events. To exploit these large amounts of data, the Supply Chain Visualizer increases supply-chain visibility by analyzing RFID data, using a mix of automated analysis techniques and human effort. The tool's core concepts include rule-based analysis techniques and a map-based representation interface. With these features, it lets users visualize the supply-chain structure, together with performance metrics, and detect problematic hot spots.

The success of VisiCalc,¹ the first interactive spreadsheet software, illustrates nicely how new technology can impact and change common work practices. Although many competing programs produced outputs that looked like ledger sheets, the dramatic – and largely unexpected – change came from the spreadsheet's visualization and interactivity. In its time, VisiCalc was a radical new idea that triggered the introduction of desktop computers to millions of offices.

RFID and other pervasive computing technologies could provide a comparable paradigm shift by empowering physical items and goods to share their identity and presence at certain locations with information systems. RFID technology can automatically generate

event data that digitally describes how physical entities, such as single items or pallets, move through supply-chain processes. Recently ratified standards, such as the EPCglobal's Electronic Product Code Information Services (EPCIS) specification,² enable a useful semantic interpretation of such RFID event data across supply chains. Because the data reflects actual business processes, we can analyze it with data mining and other techniques to detect and locate inefficiencies (such as shipment delays, inventory shrinkage, theft, and out-of-stock situations) in the physical distribution of goods.³

Considering the huge number of goods flowing through complex supply chains today and the potential impact of item-level tagging, a key issue

**Alexander Ilic,
Thomas Andersen,
and Florian Michahelles**
ETH Zurich

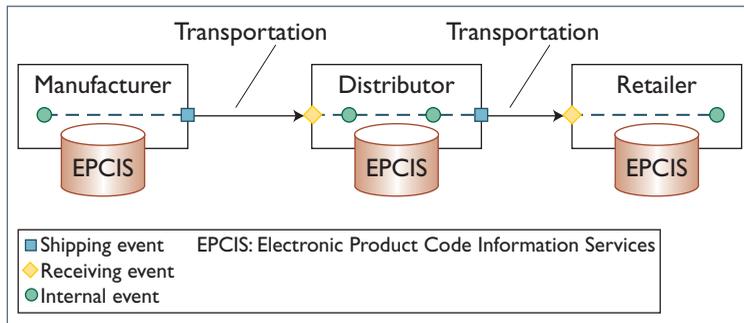


Figure 1. Generic shipping and receiving model. This event-driven representation is a sample RFID-enabled supply chain.

is understanding the supply-chain dynamics of large volumes of RFID event data. Because business users typically aren't data mining experts, we need visual representation combined with analysis frameworks to support human reasoning and perception in order to deliver useful knowledge and findings.⁴ Accordingly, we need to reduce the complexity with analysis techniques while still keeping the user in the decision-making loop.

To address these challenges, we've developed the Supply Chain Visualizer, an analysis and visualization tool for RFID event data. In this article, we present the tool, its underlying concepts, and their applications in detail. In particular, we present a concept for analyzing massive amounts of RFID event data with generic consistency rules and a map-based visualization of product-flow data combined with the detection of inconsistencies. By utilizing the semantics of RFID data and through selective aggregation, users can visualize problematic areas in a supply chain, locating sources of potential business problems. Finally, we help users explore the data by offering various time- and categorical-filtering capabilities combined with useful performance metrics and activity indices.

Supply Chain Visualizer Concept

The basic concept behind the Supply Chain Visualizer is to analyze standard RFID event data and visualize the results. (See the "Related Work in RFID Event Data Analysis" sidebar for previous work in this area.) To begin, we use a shipping and receiving model⁵ (see Figure 1) and the notion of shipping, receiving, and internal events to dynamically build up the supply-chain model. After describing the type of standard RFID event data we use, we can detail the concept of our rule-based analysis. The visualization and interaction features we've im-

plemented help present the analysis results in an intuitive way.

The EPCIS 1.0.1 specification defines the type of event data our system uses and provides a powerful and flexible way of expressing it.² Depending on the application context, different event- and master-data attributes are available. Moreover, industries as well as individual companies can extend the specification at defined extension points with custom vocabularies. Therefore, we limit ourselves to a minimum data model, a subset of the EPCIS event specification, following a what-when-where-why approach. Table 1 shows the attributes required for all of us. These events are based on the event type `ObjectEvent`, which indicates an "event pertaining to one or more physical objects identified by EPCs"² – that is, an actual RFID read.

Here's an example EPCIS event with XML binding:

```
<ObjectEvent>
  <eventTime>2008-11-01T08:24:38
</eventTime>
  <epcList>
    <epc>1234.17.327115</epc>
    <epc>1234.9.652913</epc>
  </epcList>
  <action>OBSERVE</action>
  <bizStep>urn:autoidlabs:bizstep:
    receiving</bizStep>
  <bizLocation>
    <id>urn:epc:id:sgln:0614141.33254.0
    </id>
  </bizLocation>
</ObjectEvent>
```

This example event explains that an RFID reader has read (via the action `OBSERVE`) the physical objects identified by electronic product code (EPC) numbers 1234.17.327115 and 1234.9.652913 at 8:24 a.m. on 1 November 2008. Moreover, the business step `bizStep` indicates that the objects were received (incoming shipment) by a supply-chain partner and read at a location defined by a serialized global location number (SGLN) with the uniform resource identifier (URI) `urn:epc:id:sgln:0614141.33254.0`. Systems and tools can resolve this number into a physical address and, in turn, translate it with geomapping into coordinates with the latitude and longitude values needed in the Supply Chain Visualizer for the Google maps API.

Table 1. Data attributes used for the Supply Chain Visualizer.

Attribute	Description
epcList	Contains a list of one or more electronic product code (EPC) entities, universally identifying tagged physical objects
eventTime	The time when the event occurred
action	Describes how the event relates to the life cycle of the entity specified by epcList: <ul style="list-style-type: none"> • ADD, entity created or initialized • OBSERVE, entity read and not modified • DELETE, entity destroyed or deactivated
bizStep	The business step in which this event took part; in this work, the value can be either shipping, receiving, or internal
bizLocation	The discrete business location where the tagged object is after the event occurred

To facilitate the rule-based analysis, we preprocess all incoming event data in two steps. First, we split up grouped epcList events into single events and create chronologically ordered linked lists for each individual item flow identified by the EPC. Second, we calculate time differences (based on eventTime), distances (based on the geocoordinates translation of bizLocation), and the resulting movement velocities (based on eventTime and bizLocation) for each pair of preceding and following events.

Rule-Based Analysis

The core part of our software is the analysis engine. We use generic analysis rules that can detect anomalies by evaluating individual items' preprocessed event chains. The analysis starts at each root of a product flow and checks each of the currently enabled rules for each time-ordered pair of events denoted as e_i and $e_i + 1$. Every time a pair of events fails a check, the analysis engine generates a new inconsistency event within the location of e_i .

The following rules represent checks for basic supply-chain conditions that we developed during discussions with industry experts. Because a single inconsistency violation of an individual product might not necessarily prove the existence of a bigger problem, we use aggregation techniques to illustrate hot spots based on the frequency of inconsistencies. The Supply Chain Visualizer's configuration file lets users activate or deactivate specific rules to make our tool applicable to a broad set of scenarios.

Velocity Consistency

If the event chain represents a physical item's flow, certain speed constraints apply. The speed consistency check verifies that the velocity (calculated with the eventTime and bizLocation

differences of e_i and $e_i + 1$) is between a minimum velocity v_{min} and maximum velocity v_{max} , which we define in the configuration file.

The rationale of this rule is that items can't move faster than the transportation mechanism allows. For example, if the data trace describes that a product was seen at 10 a.m. in Switzerland and an hour later in Japan, then this rule would detect that the maximum threshold v_{max} is exceeded and thus an inconsistency exists. A cloned tag (indicating a counterfeit product) might cause such an inconsistency. The default value for v_{max} is the maximum possible transportation speed of airfreight. If the transportation method for a specific product (such as a truck, ship, or plane) is known, users can more accurately configure v_{max} to reflect real-life constraints.

The minimum velocity check follows a similar rationale.

Dwell-Time Consistency

In supply chains, goods are often moved as fast as possible toward the point of sale. For instance, keeping perishable goods at a single location for too long wastes much of their valuable shelf life. Therefore, the rule of dwell-time consistency checks that the eventTime difference between e_i and $e_i + 1$ is below a threshold t_{max} . For example, inefficient stock-rotation processes might cause an inconsistency. Depending on the type of product, maximum product lifetime, or a time threshold based on a performance agreement, the configuration file must specify the maximum value t_{max} .

Life-Cycle Consistency

An EPCIS event's action field describes how an event relates to the life cycle of the physical entity being described. At the time that a physical object gets a virtual representation through the

Related Work in RFID Event Data Analysis

The geovisual analytics field is concerned with the analysis of massive amounts of movement data and their visualization so as to support decision-making for problems with spatial and temporal aspects.¹ Although the approaches for analyzing and visualizing data sets are quite advanced, they primarily rely on data with a limited set of semantic attributes — mostly only location and time. Important challenges often include extracting trips and locations from data lacking rich semantics.² In contrast, standards such as the EPCglobal's Electronic Product Code Information Services (EPCIS) enrich RFID event data with several semantic attributes of the supply-chain domain, making it easier to extract locations and routes. Furthermore, the additional semantic information offers an opportunity to develop analysis and visualization methods more specific to the supply-chain problem domain.

So far, researchers have paid little attention to analyzing and visualizing RFID event data. David Shuping and William Wright described a 3D visualization technique that illustrates object movements over time. Motivated by the context of the US Department of Defense, they applied their GeoTime technique to the RFID domain.³ Their technique drew product flows based on the geographical coordinates and combined this with events' time dimension on a z axis. Users were then able to see irregularities in the movement patterns. Although their approach works well for movements of single items, multiple items in complex topologies and the detection of problems in a supply-chain context demand another approach.

C.M. Cheung and colleagues developed a tool that can extract supply-chain topology data from RFID events by displaying supply chains as nodes and lines, which greatly reduced the complexity of RFID data.⁴ Their focus was primarily on dynamic topology changes, whereas we focus on applying rule-based analysis techniques.

I-En Liao and Wei-Chih Lin developed a model for shopping-

path analysis and data mining based on a retail store layout.⁵ Their results showed that analyzing RFID data reveals interesting patterns that let retailers better understand consumer behavior. In contrast, our work focuses on a supply-chain-level investigation. Moreover, instead of using a predefined geographic model (such as a store layout or supply-chain map), we employ a more generic approach by using EPCIS supply-chain events to extract this information.

We aim to advance the existing body of knowledge in several ways. First, we combine the concept of geographical visualizations with rule-based analysis techniques of the specific problem domain of RFID event data. Second, we show that an analysis based on generic consistency rules can deliver useful results for supply-chain managers. Third, we use a supply-chain model derived from data, making our solution applicable to a wide area of interorganizational contexts. Finally, we provide an explorative map-based visualization interface that combines supply-chain topology data with the output of the rule-based analysis.

References

1. G. Andrienko et al., "Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda," *Int'l J. Geographical Information Science*, vol. 21, Jan. 2007, pp. 839–857.
2. G. Andrienko, N. Andrienko, and S. Wrobel, "Visual Analytics Tools for Analysis of Movement Data," *ACM SIGKDD Explorations Newsletter*, vol. 9, no. 2, 2007, pp. 38–46.
3. D. Shuping and W. Wright, "Geotime Visualization of RFID," *RFID J.*, Mar./Apr. 2005, pp. 1–6.
4. C.M. Cheung et al., "An RFID Enabled Topology Visualization System for Supply Chain Management," *Ann. J. IIE*, vol. 26, 2006, pp. 61–70; www.ise.polyu.edu.hk/output&pub/output/0506web.htm.
5. I.-E. Liao, and W.-C. Lin, "Shopping Path Analysis and Transaction Mining Based on RFID Technology," *Proc. 1st Ann. RFID Eurasia*, IEEE Press, 2007.

EPC, the RFID reader generates an event with `action=ADD`. Once the tag is deactivated or the physical object is destroyed, the RFID reader then creates an event with `action=DELETE`.

All standard RFID reads between the object's creation and disposal trigger `action=OBSERVE` events. The life-cycle consistency rule therefore checks that there are no events before an ADD action event and no events after a DELETE action event. An inconsistency might indicate a counterfeit product or that a reader or middleware has been improperly configured.

Pair-Wise Shipping and Receiving Confirmation

Some industries with strong regulations on

traceability, such as food and pharmaceuticals, widely use pair-wise shipping and receiving confirmations. In some cases, the chain of custody is even more explicitly documented using signed electronic pedigree documents. This rule checks that for every $e_i + 1$ with `bizStep=receiving`, there must be a corresponding e_i with `bizStep=shipping`. Moreover, an event e_i with `bizStep=shipping` can only be followed by an event $e_i + 1$ with `bizStep=internal` or `bizStep=receiving` to detect inconsistencies with duplicate send and receive business steps.

Transition Probability Consistency

The transition probability check is based on additional reference data, which is precalculated

during the start of the Supply Chain Visualizer. Depending on the configuration file setting, for example, the system uses the first 10,000 events as a training set to compute transition probabilities from one supply-chain location to another for each product group. During the analysis phase, we calculate the transition probabilities by counting the total product flows and then compare them to the training set's probabilities. The consistency rule checks that the difference between the training and analyzed data sets doesn't exceed a predefined threshold.

In contrast to the previous rules, the threshold check is performed against data groups of supply-chain paths rather than against single items. The default setting allows a maximum deviation of 5 percent from the training set. An inconsistency indicates, for example, a change in route utilization vis-à-vis the training data set.

Visualization and User Interface

The primary role of the system's visualization component is to reduce the complexity of the underlying data and to equip a user with powerful filtering techniques. The following sections present the design rationales and concepts for this component.

Map-Based View

By interviewing several supply-chain experts, we learned that a geographic representation of the supply chain is easy to understand and communicate. A key advantage for users is that they can zoom in and pan over to get a more detailed view on a specific area or zoom out to get a broader, higher-level perspective of the supply chain.

We kept the tool's visualization elements simple to help users see the supply chain's status at a glance (see Figure 2). Circle-shaped node icons identify the supply-chain partners' locations. Lines represent product flows and connect the nodes with each other. Instead of drawing each single product-flow line on the map, the system aggregates all analyzed data to draw just a single line to connect two nodes. While aggregating the data, the analysis engine calculates the product flow direction and frequencies. Accordingly, each line contains one or two directions, and the line's thickness represents a particular link's product flow frequency. Similar to subway maps, our interface lets users easily see the supply chain's main routes.



Figure 2. The Supply Chain Visualizer user interface. Green means no inconsistencies, yellow means that there are some inconsistencies but the number is below a predefined threshold, and red means that the number of inconsistencies at the particular node exceeds the threshold and thus indicates a serious problem.

Based on the output of the analysis, we use a simple three-color coding scheme to show users a node's status. We use the same color scheme to indicate when the transition probability analysis rule found deviations from the training data set.

Performance Metrics

The system interface also shows users useful supply-chain performance metrics. When a user moves the mouse over a supply-chain node, the numbers of incoming, outgoing, and remaining items are displayed together with their average dwell time. When moving the mouse over a connecting line's direction arrow, the interface displays the total number of items that moved in this direction (total product flow) and the average transport time in a tool-tip window (see Figure 3).

Data Selection and Filtering

The data selection and filtering component changes the data basis for the analysis and thus updates the map-based view with Ajax calls. By using the EPC's hierarchical structure, we can filter the analyzed data set by product manufacturer (or tag owner), product class, and serialized product ranges. Moreover, we provide an

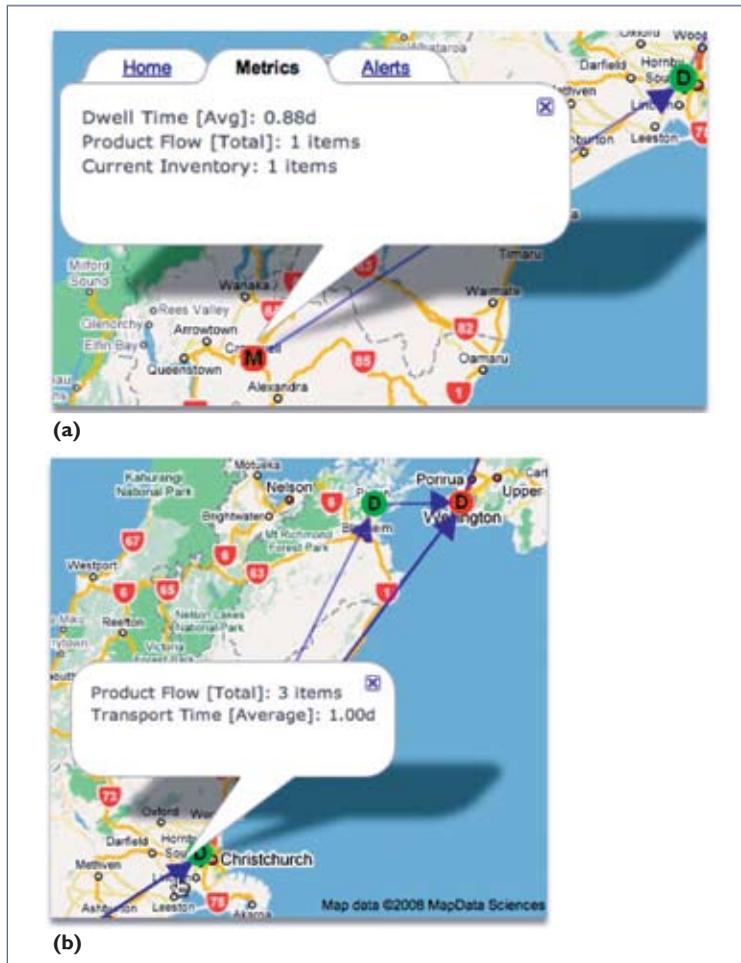


Figure 3. Performance metrics. The Supply Chain Visualizer automatically calculated performance metrics of (a) nodes and (b) routes.

event activity chart, which shows the frequency of events against the time axis.

With the combination of these two features, a user can conveniently drill down the data into details to find out specific products and time ranges for detected inconsistencies.

Advanced Application Scenarios

Several application scenarios illustrate how users can track down complex business problems with the Supply Chain Visualizer. Although the following scenarios are based on sample data (instead of data from real-life RFID implementations), we discussed them in detail with our industry partners, who identified them as a high priority. Moreover, these application scenarios have been demonstrated with our prototype at an academic conference.⁶ For each scenario, we illustrate a business problem and how the Supply Chain Visualizer can visualize the source

of the problem. However, we must interpret the data carefully because several business problems can reveal themselves with similar evidence. With the Supply Chain Visualizer, we mitigate the problem of false alarms using helpers such as the filtering techniques, performance metrics, and flexible rule configuration mechanisms. Thus, we keep the human in the decision-making loop to interpret the data and identify the actual source of the problem.

Theft

In the US, the retail industry estimates it loses at least \$25 billion each year due to theft from distribution channels and stores.⁷ In this context, let's consider a scenario in which a truck transports goods from a supplier to a retailer. A malicious actor might be able to steal goods from the truck at several occasions. The critical phases occur primarily after loading a truck at the supplier, before unloading a truck at the retailer, and during a driving break.

Because the number of outgoing items at the supplier's end must match the number of incoming items at the retailer's end, the Supply Chain Visualizer can detect such problems by utilizing the pair-wise shipping and receiving consistency check. For example, if an item gets stolen during transport, the shipping event at the supplier's end has no corresponding receiving event at the retailer's end. The system will detect single missing items, and a supply-chain manager can browse and explore the patterns of such inconsistencies to develop suitable countermeasures.

Counterfeiting

European customs seize up to 100 million counterfeit and pirated goods every year. Counterfeiting isn't just a huge loss of revenue and research and development investments for manufacturers and brand owners, it's a severe threat to consumer safety.

An important problem in this context is the cloning of security features. In the case of RFID, this could mean the cloning of unique product numbers. When fake products with copied identifiers are injected into the legitimate supply chain, they also produce valid RFID events. With the transition probability check, we can determine whether a set of events, having the same identifier, was created by a single product or by two products (genuine and counterfeit).

Thus, the Supply Chain Visualizer can detect such inconsistencies by highlighting improbable transitions (see Figure 4), stimulating location-based product authentication.⁸

Inventory Shrinkage

Inventory shrinkage due to spoilage is a major problem for the food industry. More than 10 percent of the total commercial waste in the UK is caused by spoiled food products, which can be attributed to distribution failures. The key factors here include variations in environmental parameters and unnecessary time spent on vehicles or in distribution centers.⁹

With the dwell-time consistency check, the Supply Chain Visualizer detects whether a product spent more time at a specific location than intended. Because our analysis is based on multiple product flows, the color scheme will show problematic areas as soon as the consistency violations accumulate. In addition, the automatic performance metrics let the user quickly see where a product spent most of its lifetime. Consequently, users can detect patterns over time and thus can find potential sources of spoilage.

Implementation

Different supply-chain partners often want to store event data in their own EPCIS repositories, where they can manage the access control, rather than in a centralized database. Therefore, the Supply Chain Visualizer's technical architecture (see Figure 5) must connect to the different EPCIS systems and retrieve the RFID event data. So far, we've successfully connected the Supply Chain Visualizer with EPCIS systems running the FossTrak EPCIS.¹⁰ Involved organizations can establish a connection between the Supply Chain Visualizer and their information systems on demand. This means that the data collection and analysis can be run only with the consent of each of the EPCIS repository owners. Thus, we assume that they primarily use the Supply Chain Visualizer for strategic evaluations, possibly once a month rather than daily.

The Supply Chain Visualizer itself consists of four main components (see Figure 5):

- The *trace data collector* accesses the remote RFID event repositories in a standardized way (as defined by the EPCIS interface spec-



Figure 4. User interface. Here, we can see the clone-detection feature based on transition probability checks.

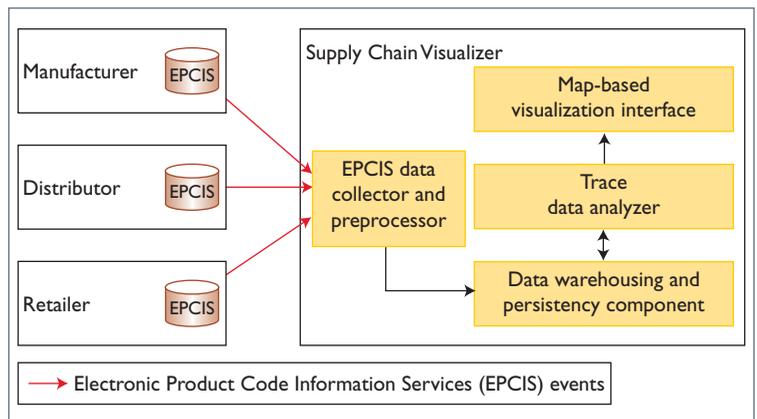


Figure 5. Architecture of the Supply Chain Visualizer and the data flows between the components. The boxes represent the components of the Supply Chain Visualizer, and the arrows represent the data flow.

ification²). It collects the EPCIS event data from different supply-chain partners and then harmonizes, preprocesses, and stores it in a data warehouse database.

- The *data warehouse*, the basis for all analysis operations, contains both the input data from the trace data collector and the intermediate results from the trace data analyzer components.
- The *trace data analyzer* contains the rule-based analysis techniques that we described earlier. This component interacts with the data warehouse and creates aggregated output data for the user interface.
- The GUI component provides a map-based visualization interface, which displays the

results from the trace data analyzer and interacts with the user to receive filtering and configuration inputs.

Our research results open up several opportunities for future research. First, our approach is based on a classical data warehouse analysis approach. Although this analysis method was completely sufficient to demonstrate the concepts, in principle, RFID also enables a real-time approach. The generic consistency rules we proposed here could be applied in a real-time analytics system and could process streams of RFID data to trigger alerts or notifications as soon as inconsistencies occur. Second, we could also add more user guidance and workflow support to the system to provide an executive summary with recommended actions and probable high-level business issues. Third, it's possible to add extra business value and get better analysis results by considering additional sensor data such as temperature, humidity, light, or shock in the generic consistency rules. This extension would prove especially useful in pharmaceutical or food supply chains. Lastly, we could connect external data sources such as enterprise resource planning systems to the analysis engine to detect differences between physical observations (approximated through automated identification and tracking with RFID) and, for example, expected stock levels based on transactional data of other sources (such as electronic date interchange or human input).

We're also looking forward to exploring the applicability of our tool in different industries. Particularly, once big players such as Metro and Wal-Mart share their data across the supply chain, it would be interesting to evaluate our tool in a real-world scenario. 

Acknowledgments

This work was partly funded by the European Commission within the Sixth Framework Program (2002–2006) IP number IST-FP6-033546.

References

1. S. Gal, "Footholds for Design," *Bringing Design to Software*, ACM Press, 1996, pp. 215–231.
2. *EPCIS 1.0.1 Specification*, EPCglobal, 2007; www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf.
3. H.L. Lee, and O. Ozer, "Unlocking the Value of RFID," working paper, Graduate School of Business, Stanford Univ., 2005.
4. J.J. Thomas, and K.A. Cook, "A Visual Analytics Agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, 2006, pp. 10–13.
5. D. McFarlane, and Y. Sheffi, "The Impact of Automatic Identification on Supply Chain Operations," *Int'l J. Logistics Management*, vol. 14, no. 1, 2003, pp. 1–17.
6. A. Ilic et al., "Analyzing Product Flows with the Supply Chain Visualizer," demo at Int'l Conf. for Industry and Academia (Internet of Things 2008), 2008; www.im.ethz.ch/publications/Demo_SCVIZ_IOT2008.pdf.
7. R. Koh et al., "Prediction, Detection, and Proof: An Integrated Auto-ID Solution to Retail Theft," white paper series, MIT-AUTOID-WHO22, Auto-ID Labs, 2003.
8. M. Lehtonen, F. Michahelles, and E. Fleisch, "Probabilistic Approach for Location-Based Authentication," *Proc. 1st Int'l Workshop Security for Spontaneous Interaction (IWSSI)*, Springer, 2007, pp. 486–491.
9. A. Osvald, and L.Z. Stirn, "A Vehicle Routing Algorithm for the Distribution of Fresh Vegetables and Similar Perishable Food," *J. Food Eng.*, vol. 85, no. 2, 2008, pp. 285–295.
10. C. Floerkemeier, C. Roduner, and M. Lampe, "RFID Application Development with the Accada Middleware Platform," *IEEE Systems J.*, vol. 1, no. 2, 2007, pp. 82–94.

Alexander Ilic is a PhD candidate in the Department of Management, Technology, and Economics at ETH Zurich and works as senior researcher at the Bits to Energy Lab and the Auto-ID Labs Zurich/University of St. Gallen. His research interests include sensor-enhanced supply-chain operations and their efficiency optimizations. Ilic has an MS in computer science and computational linguistics from the Technical University of Munich. He is a member of the IEEE. Contact him at ailic@ethz.ch.

Thomas Andersen is a telecommunication engineer at Muzui in Copenhagen. His research interests include the analysis of RFID trace data. Andersen has an MS in telecommunications from the Technical University of Denmark. Contact him at thomas@andersen.net.

Florian Michahelles is the associate director of the Auto-ID Labs Zurich/St. Gallen and manager of the information management lab of Professor Elgar Fleisch at ETH Zurich. His research interests include topics centered around RFID applications, EPC network, anticounterfeiting, and mobile RFID applications for the consumer. Michahelles has a PhD from ETH Zurich for his research in participative design of wearable computing applications and the development of innovative business cases for ubiquitous computing. Contact him at fmichahelles@ethz.ch.