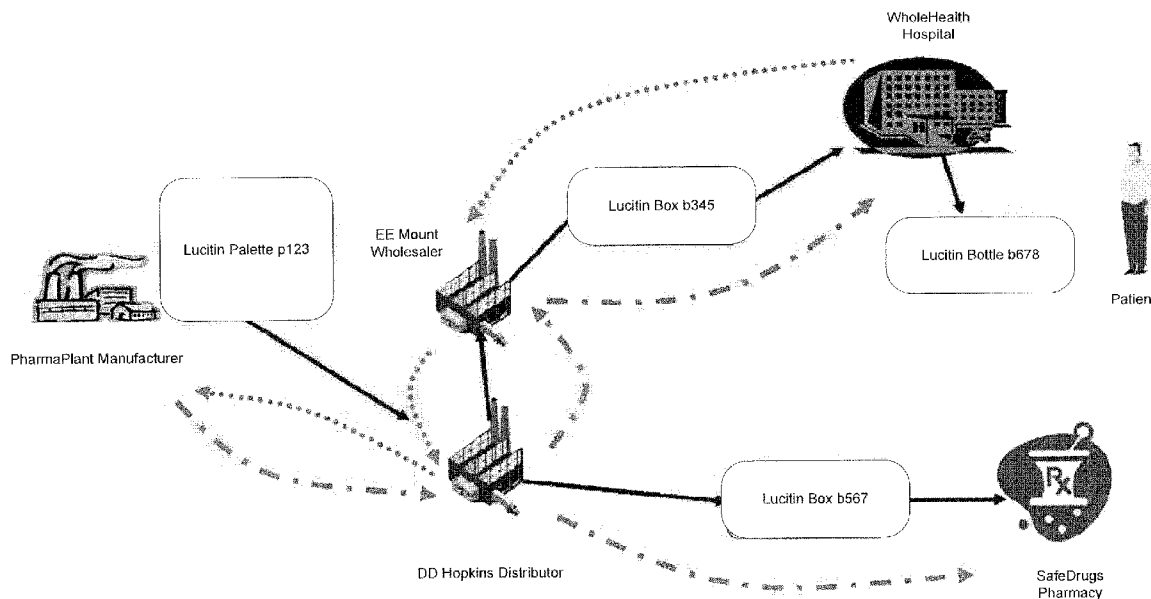




US 20080129452A1

(19) **United States**(12) **Patent Application Publication**
AGRAWAL et al.(10) **Pub. No.: US 2008/0129452 A1**(43) **Pub. Date: Jun. 5, 2008**(54) **MIDDLEWARE FOR QUERY PROCESSING
ACROSS A NETWORK OF RFID DATABASES**(22) Filed: **Dec. 5, 2006**(75) Inventors: **RAKESH AGRAWAL, SAN
JOSE, CA (US); ALVIN K.
CHEUNG, SAN FRANCISCO, CA
(US); KARIN KAILING, SAN
JOSE, CA (US); STEFAN
SCHOENAUER, SAN JOSE, CA
(US)****Publication Classification**(51) **Int. Cl.**
H04Q 3/02 (2006.01)
(52) **U.S. Cl.** **340/10.1**
(57) **ABSTRACT**Correspondence Address:
IP AUTHORITY, LLC
RAMRAJ SOUNDARARAJAN
4821A Eisenhower Ave
Alexandria, VA 22304(73) Assignee: **INTERNATIONAL BUSINESS
MACHINES CORPORATION,**
ARMONK, NY (US)(21) Appl. No.: **11/566,931**

An implementation wherein RFID data is shared across independent organizations has been addressed. RFID data is usually spread across different parties, e.g. enterprises in a supply chain and thus, efficient query processing across all parties is required. Traceability is emerging as one of the key applications of RFID technology. A generic data model is introduced for querying RFID data across a network of independently operated data sources. The model can be used to facilitate traceability query processing and give a set of representative traceability queries. A newly designed process-and-forward approach is implemented for executing traceability queries.



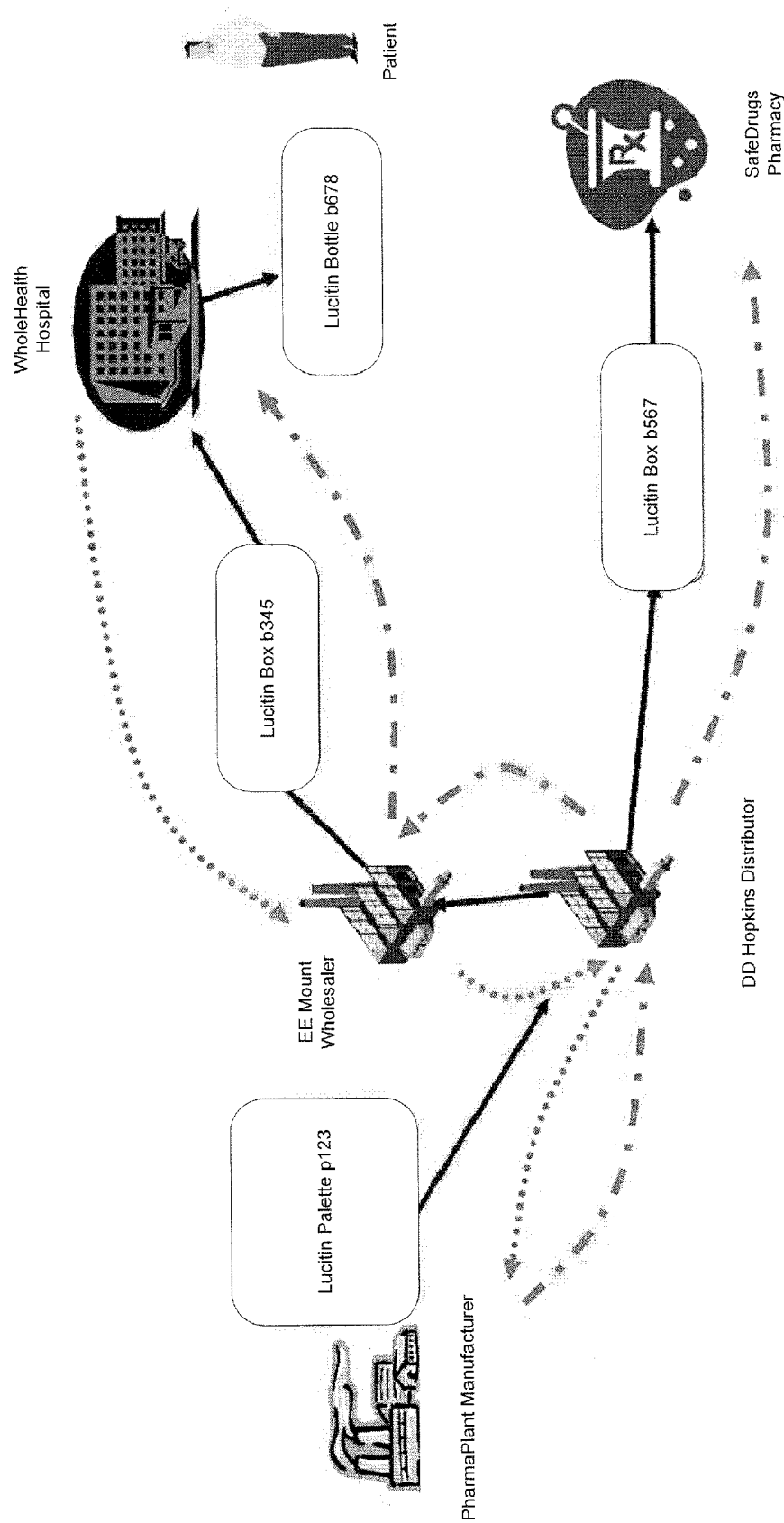


FIGURE 1

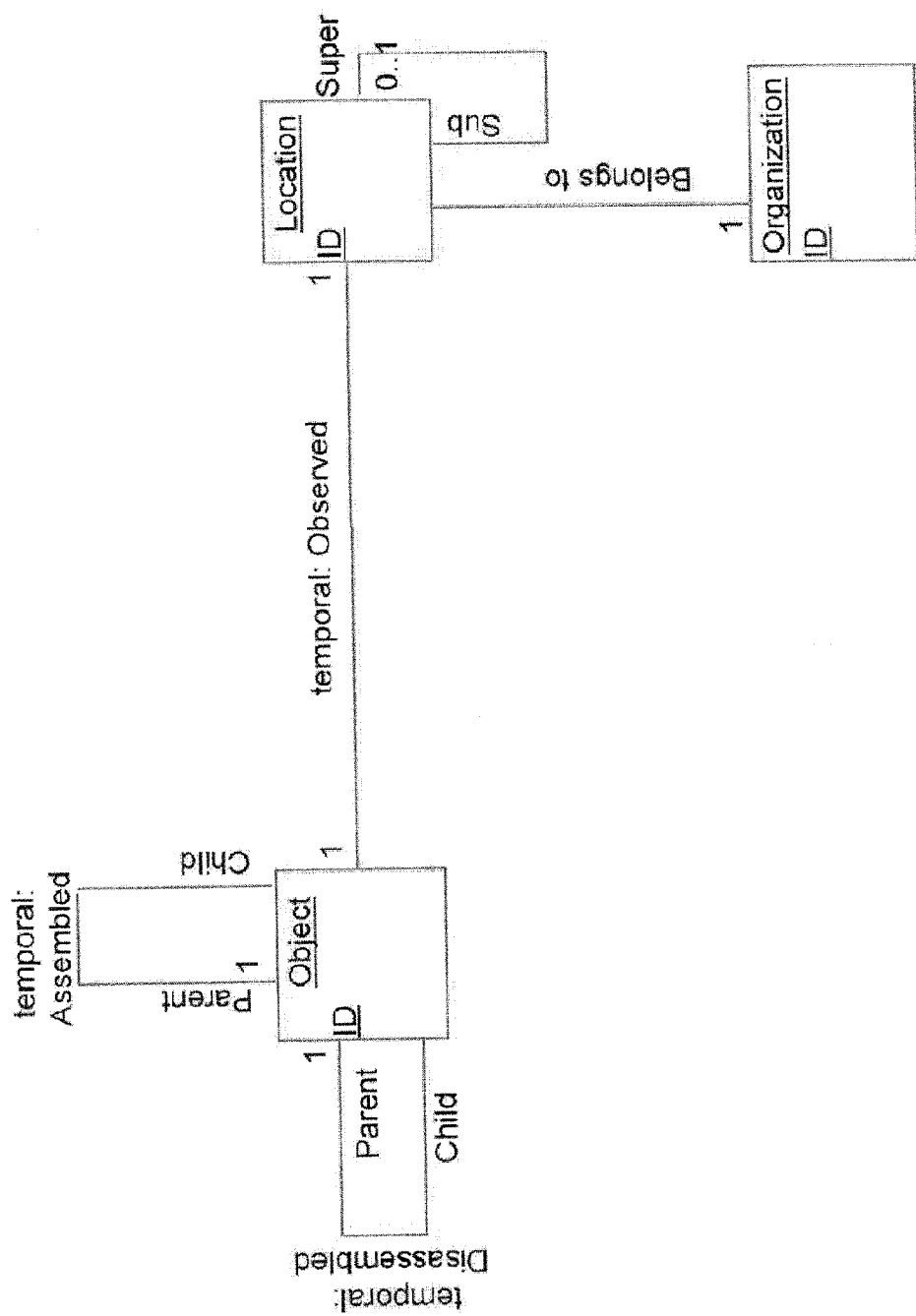


FIGURE 2

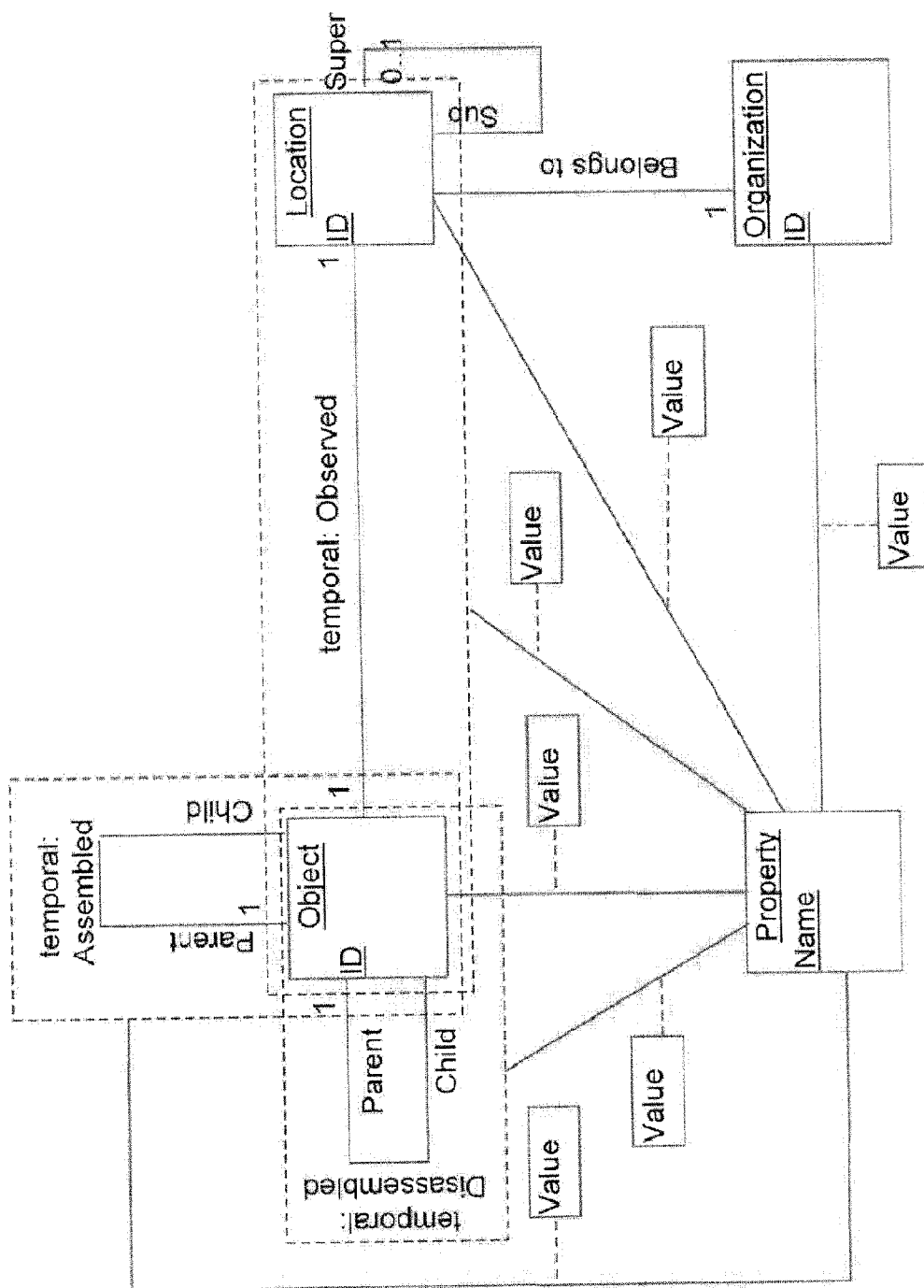


FIGURE 3

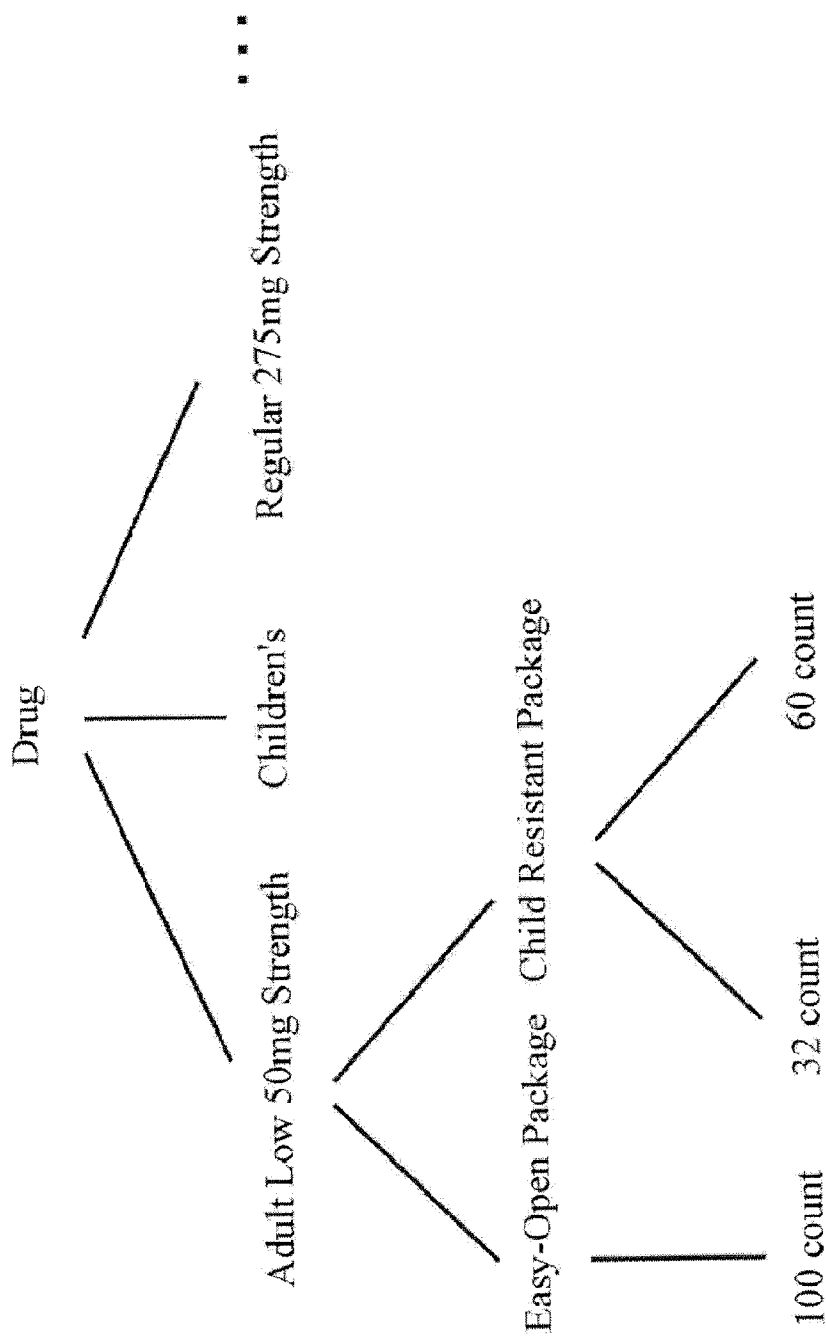


FIGURE 4

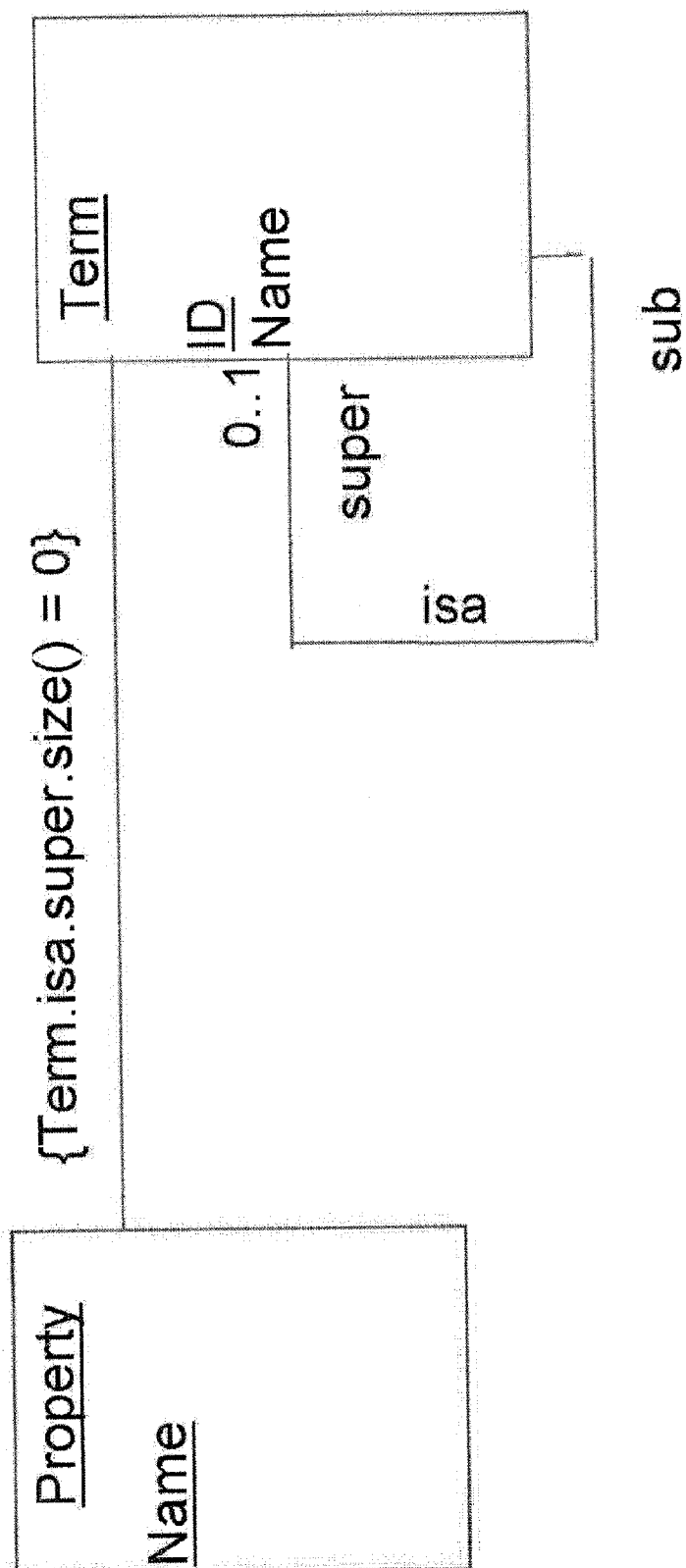


FIGURE 5

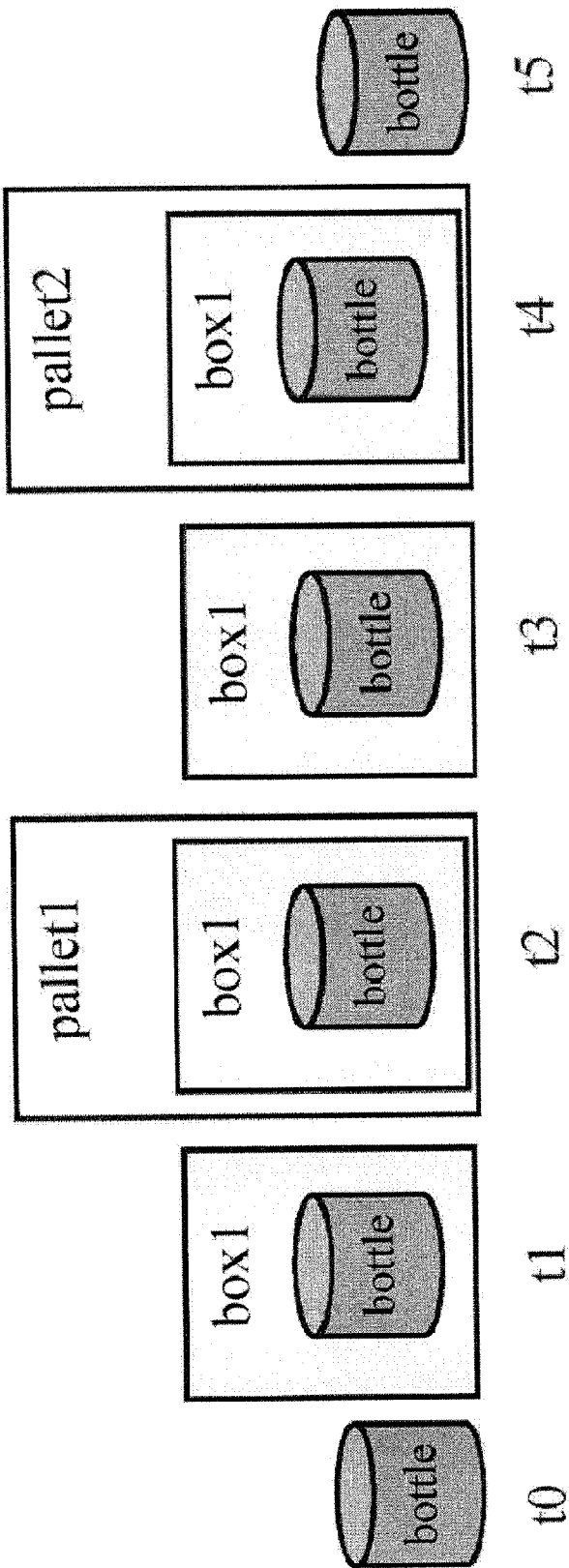


FIGURE 6

Observed		ObsPropertySet		Assembled	
oid	lid ts	... propertyName	value	parent	child ts
CYM	cylinder1 11 t1	sentTo	ENM		
	cylinder2 11 t1	sentTo	ENM		
ENM	cylinder1 15 t5	receivedFrom	CYM	engine1	cylinder1 t10
	cylinder2 15 t5	receivedFrom	CYM	engine1	cylinder2 t10
	cylinder1 10 t10				
	cylinder2 10 t10				
CAM	engine1 10 t10				
	engine1 15 t15	sentTo	CAM		
	engine1 20 t20	receivedFrom	ENM	car1	engine1 t25
	engine1 22 t25				
	car1 22 t25				

FIGURE 7

MIDDLEWARE FOR QUERY PROCESSING ACROSS A NETWORK OF RFID DATABASES

BACKGROUND OF THE INVENTION

[0001] 1. Field of Invention

[0002] The present invention relates generally to the field of middleware. More specifically, the present invention is related to middleware for query processing across a network of RFID databases.

[0003] 2. Discussion of Prior Art

[0004] In recent years, Radio-Frequency Identification (RFID) has attracted a lot of attention. Although RFID technology is not new and can be traced back to World War II, a number of recent developments have accelerated the adoption of RFID technology in different industries.

[0005] Advancements in RFID physics and hardware technology have been pushing the average price of individual passive tags to be lower than ever. An initiative is underway in Japan to produce a 5 yen RFID tag by the end of 2006, 1.8 billion tags have been sold in 2005, and the total market opportunity for RFID has been predicted to reach seven billion dollars in 2008.

[0006] A number of legislations have pushed industries to consider RFID technology for compliance purposes. Recent laws passed in the US will require pharmaceutical industry to provide a valid pedigree of drug items upon request. Similar legislations have been formulated in the food industry, such as the Japanese Beef Traceability Law, the US Department of Agriculture national animal identification system, and the EU requirements on fish and fish products traceability.

[0007] Standardization efforts by industry consortia such as EPCglobal™ (formerly Auto-ID Center) have further promoted the adoption of RFID technology. EPCglobal™ is creating standards for RFID data communication such as a specification for RFID tag numbering formats, a transmission protocol to obtain information from RFID readers, and an overall architecture for a network of databases containing RFID data.

[0008] Recently several industry research papers have been published on RFID data management. SAP presented an overview on their existing RFID infrastructure in the paper titled “Integrating Automatic Data Acquisition with Business Processes—Experiences with SAP’s Auto-ID infrastructure”. Under “Lessons Learned”, they state that companies need to overcome their reluctance to collaborate because the full potential of RFID technology can only be unlocked through collaboration and data sharing across sites and organizations. ORACLE™ presented a new bitmap data type for ORACLE DBMS to support RFID-based item tracking applications in the paper titled “Supporting RFID-based Item Tracking Applications in ORACLE DBMS”. Siemens proposed a temporal data model for their RFID data management system in the paper titled “Temporal Management of RFID data.” OAT Systems™ gave a brief introduction to RFID technology, highlighting some of the data management challenges in the paper titled “Managing RFID Data”.

[0009] An RFID cube is introduced in the paper titled “Warehousing and Analyzing Massive RFID Data Sets” to support warehousing and analysis of massive RFID data sets. Apart from this work, academic research has mostly focused on privacy and security issues surrounding communication between RFID reader and RFID tag. An overview can be found in the paper titled “Radio Frequency Identification: Adversary Model and Attacks on Existing Protocols”. How-

ever, confidentiality of RFID data once it is stored in databases is not addressed. None of these papers presents solutions to the challenges imposed by independent organizations sharing data.

[0010] In the area of federated database systems, the issue of querying across heterogeneous data sources has been addressed, but those solutions rely on a priori knowledge about the data distribution (see paper to Kossmann titled “The State of the Art in Distributed Query Processing”). But with RFID traceability systems the distribution is unknown, since the tracked objects can move freely between organizations.

[0011] For peer-to-peer databases, work has been done on locating documents, which means to find the single place where all the information about an object is stored (see paper to Androutsellis-Theitokis et al. titled “A Survey of Peer-To-Peer Content Distribution Technologies” and the paper to Stoica et al. titled “A Scalable Peer-To-Peer Lookup Service for Internet Applications”). But in traceability systems the information about an entity is spread over several participating databases and the set of those databases may change at any point in time. Some ideas on collecting information about a single entity from several databases in a peer-to-peer setting have been presented in the paper to Giunchiglia et al. titled “Making Peer-to-Peer Databases Interact—A Vision for an Architecture Supporting Data Coordination”.

[0012] In the following, two existing industry solutions for implementing query processing in traceability networks are examined. The architectures differ in the amount of data-distribution supported. A central warehouse solution where every organization publishes its RFID data to a central site is first described. After that, a solution that is proposed by EPCglobal™ is described where each organization keeps RFID data in a local repository and only publishes data to central directory services (see paper to Chawathe et al. titled “Managing RFID Data” and paper to Traub et al. titled “The EPC-global Architecture Framework”).

[0013] In the data warehouse approach, RFID data collected within each organization is published to a central data warehouse. In this case, all organizations have to agree on a common storage format for RFID data as well as for all property data they want to share with each other. Together with their data, each organization also has to publish its confidentiality or data sharing policy to the central data warehouse. Mechanisms such as web services are provided to organizations to query the stored data based on the policies installed, and query processing is performed entirely in the warehouse.

[0014] Since the heterogeneity aspects (e.g. data schema differences) do not exist in this approach, query processing is simplified as all data can be accessed in a uniform way. It becomes possible to do optimizations such as the RFID cube proposed in the paper to Gonzalez et al. titled “Warehousing and Analyzing Massive RFID Data Sets”, and incoming queries can be executed as-is against the database. However, as a query might span data from multiple owners there needs to be a way to detect and enforce multiple policies from different organizations. Additionally, as the amount of RFID data increases, the total amount of data that needs to be published may put serious constraints on such a central warehouse approach.

[0015] Rather than sending all data to a central warehouse, an alternative would be to allow data to be stored in local repositories at each organization and make those repositories accessible in the traceability network. The most notable pro-

posals in this regard is the EPCglobal Framework (see paper to Chawathe et al. titled "Managing RFID Data" and paper to Traub et al. titled "The EPCglobal Architecture Framework"), which consists of a network with nodes (referred to as subscribers in EPCglobal Framework), and a number of central registries (called core services) that the nodes can utilize. Each node offers a simple, standardized query interface (called information service) to a repository with RFID data. An application (called accessing application) can use the standardized query interface of a repository in order to obtain data.

[0016] The challenge in such a network of distributed repositories is, given a traceability query, to locate the data sources that contain tuples that contribute to the answer construction. Central directory tables can be used to guide a query to the necessary nodes. The EPCglobal Framework proposes an Object Naming Service (ONS) and a Discovery Service (DS) as its core services. The Object Naming Service provides a centralized registry through which an object may be associated with the information service at the node where the object or more specifically its tag was created. An application may also use the Discovery Service to locate the information service of all EPCglobal subscribers that have information about the object in question. This ensures that even if the other EPCglobal subscribers within a supply chain are not known to an application, it will be able to locate all information concerning a specific object.

[0017] All nodes have to update the core services with relevant information, for instance register with the Object Naming Service when a new RFID tag is created, or update the Discovery Service when a tag moves from one node to another.

[0018] Whatever the precise merits, features, and advantages of the above mentioned prior art techniques, none of them achieves or fulfills the purposes of the present invention.

SUMMARY OF THE INVENTION

[0019] In one non-limiting example, a network of independently operated nodes sharing REID data comprises: (a) a local RFID repository; and (b) a query engine, wherein the query engine (implemented, for example, using middleware) receiving a query, retrieving data related to the query from the local RFID repository, analyzing, executing, and optionally rewriting the query based on retrieved data, and forwarding said rewritten query to another of the independently operated nodes, wherein the query is efficiently processed by combining results of query executions across a plurality of independent nodes.

[0020] In another non-limiting example, a method implemented as middleware in a node among said network of independently operated nodes comprises the steps of: (a) receiving a query; (b) retrieving data related to said query from a local RFID repository; (c) analyzing, executing and, optionally, rewriting said query based on retrieved data; and (d) forwarding said rewritten query to another of said independently operated node, wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

[0021] In yet another non-limiting example, an article of manufacture comprising a computer user medium having computer readable program code embodied therein implementing a middleware to query RFID data across a network of independently operated nodes sharing RFID data, wherein said middleware comprises: (a) computer readable program

code aiding in receiving a query; (b) computer readable program code aiding in retrieving data related to said query from a local RFID repository; (c) computer readable program code analyzing, executing and, optionally, rewriting said query based on retrieved data; and (d) computer readable program code aiding in forwarding said rewritten query to another of said independently operated node, wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

[0022] In yet another non-limiting example, a method implemented across a network of independently operated nodes sharing RFID data comprises the steps of: (a) receiving a query; (b) analyzing and rewriting the query; (c) retrieving data related to the rewritten query from a local RFID repository and evaluating local results; (d) identifying additional nodes for forwarding the rewritten query; (e) forwarding the rewritten query to said identified nodes, where said rewritten query is evaluated using data in corresponding local RFID repositories to produce remote results; (f) obtaining results of evaluated query from said identified nodes; and (g) combining said evaluated local results with said remote results, wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] FIG. 1 illustrates traceability queries in a supply chain.

[0024] FIG. 2 illustrates an ER diagram of the present invention's basic model.

[0025] FIG. 3 illustrates an ER diagram of the present inventions complete model.

[0026] FIG. 4 illustrates part of the GTIN taxonomy.

[0027] FIG. 5 illustrates the present invention's integration with a taxonomy.

[0028] FIG. 6 illustrates the packaging history of a bottle.

[0029] FIG. 7 illustrates example data produced by three organizations CYM, ENM, and CAM.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0030] While this invention is illustrated and described in a preferred embodiment, the device may be produced in many different configurations, forms and materials. There is depicted in the drawings, and will herein be described in detail, a preferred embodiment of the invention, with the understanding that the present disclosure is to be considered as an exemplification of the principles of the invention and the associated functional specifications for its construction and is not intended to limit the invention to the embodiment illustrated. Those skilled in the art will envision many other possible variations within the scope of the present invention.

[0031] Traceability is emerging as the key application of RFID technology. Three prototypical examples are provided below of how RFID data can be used to trace products in a supply chain (see FIG. 1 for an illustration).

[0032] Pedigree Generation

[0033] At WholeHealth hospital a patient gets treated with a drug called Lucitin and shows an adverse reaction. The hospital starts an investigation to find out if something went wrong with the drug. If the drug bottle contains an RFID tag and its movement throughout the supply chain was recorded, the investigator will be able to automatically create the pedi-

gree of the drug by tracing back its steps in the supply chain (see dotted lines in FIG. 1). Starting with the hospital each enterprise that has handled the drug will provide information, such as shipment dates and employees who processed the drug. The pedigree document can then be used to verify that the drug bottle under investigation indeed came from the manufacturer that is stated on the bottle and is not a counterfeit product.

[0034] Targeted Recall

[0035] Upon investigation PharmaPlant detects that not only bottle b678 but the whole pallet p123 containing the bottle is unsafe and needs to be taken out from the market. Traceability data allows the manufacturer to trace the forward path of p123 and its content throughout the supply chain to locate precisely the bottles that must be recalled (see dashed lines in FIG. 1), without having to issue an indiscriminate recall of every single bottle of Lucitin on the market.

[0036] Problem Detection

[0037] Traceability data can also be used to discover unusual patterns within the supply chain. For example once a set of defective products was found, an investigation might show that the majority of defective products was produced by a certain assembly line, or that most counterfeited products were handled by a certain distributor.

[0038] There are some challenges to traceability data management that must be addressed in order to realize the applications described.

[0039] Sharing Across Independent Organizations.

[0040] In most traceability applications data is spread across independent enterprises. For example, to answer recall or pedigree queries, all involved enterprises need to be located and data from heterogeneous data sources needs to be retrieved. Since data is owned by different, potentially competing, enterprises, mechanisms to limit data disclosure are necessary.

[0041] Containment of Objects.

[0042] As products move between enterprises, they can be packed into containers (e.g. pallets or boxes). In the example scenario, PharmaPlant sent a pallet of Lucitin to the distributor DDHopkins, where the pallet was broken down into two boxes. One of the boxes was sent to WholeHealth hospital via a wholesaler. At the hospital the individual pill bottles got unpacked. While the initial request for a pedigree at the hospital is for pill bottle b678, as the pill bottle's path is traced back and the request is forwarded to the distributor, the request will need to be changed to box b345 and subsequently to pallet p123.

[0043] Scalability.

[0044] The volume of data generated as a consequence of having RFID-tagged objects is enormous. A major electronics manufacturer estimates that the RFID implementations at three of its manufacturing plants alone will generate between 1 and 5 Terabytes of data per day (see literature titled "RFID: The Complete Guide"). Additionally the number of organizations that to adopt RFID technology to enable traceability is growing rapidly.

[0045] The present invention provides a system design for data management in RFID networks, geared towards efficiently and effectively supporting traceability applications. The present invention's design provides a solution to the challenges outlined above. The focus is on enabling 15 applications that depend on sharing traceability data across independent enterprises.

[0046] It is assumed that RFID data is already captured, cleaned, and stored in one or more databases. Issues related to synchronizing clocks are not addressed. Time synchronization has been studied in sensor networks (e.g. see the paper to Sunararaman et al. titled "Clock Synchronization for Wireless Sensor Networks: A Survey") as well as in ad-hoc networks (e.g. see paper to Romer et al. titled "Time Synchronization in Ad Hoc Networks"). While the requirements for time synchronization within an organization can be high, inter-organizational time synchronization is often not an issue for traceability applications. There is usually a sufficient time gap between observations made at two different organizations to tolerate asynchrony.

[0047] When designing a conceptual model for traceability networks:

[0048] The model should capture common aspects of traceability networks while allowing extensions to address specific requirements. For example, each traceability network requires objects to have a unique identifier but only pharmaceutical supply chains require that objects additionally have a national drug code.

[0049] The model should reflect the fact that independent organizations interact. It should reduce data dependencies between organizations as much as possible.

[0050] The model should provide participants of a traceability network with the ability to specify queries across the entire network. It should allow the formulation of a query without knowledge about how the data is stored, where it is located, and how a query is executed.

[0051] Basic Traceability Model

[0052] The traceability model is based on an in depth study of a number of different traceability applications including supply chain optimization, inventory visibility, pedigree generation, product recall, patient surveillance in a hospital, and cargo tracing for homeland security. ER constructs (e.g., see paper to Chen titled "The Entity-Relationship Model - Toward a Unified View of Data") are used to describe it.

[0053] Three entity types are derived: object, location, and organization.

[0054] Objects. Objects are real world entities that can be traced. Products that move through a supply chain are typical examples of objects. Objects are uniquely identifiable, for example using the Electronic Product Code (EPC).

[0055] Locations. A location is any place where objects are observed. Example locations are "manufacturing plant conveyor belt #1" and "distribution center dock door #2". Each location within a traceability network has to be uniquely identifiable, for example by a Global Location Number (GLN). Locations can be hierarchically organized as is the case for GLNs.

[0056] Organizations. An organization (e.g. manufacturer, distributor, hospital) is responsible for a set of locations. Organizations have to be uniquely identifiable. The UCC Company Prefix, a globally unique number assigned to member companies of the UCC is a common identification scheme for organizations.

[0057] Objects interact with their environment and with each other and thereby generate events. Three different event types are presented: observation, assembly, and disassembly events, denoted as three different relationships Observed, Assembled, and Disassembled in the ER diagram (see FIG. 2).

[0058] Observed. The relationship Observed captures the fact that an object was seen at a certain location at a certain

time. A ternary relationship exists between object, location, and a timestamp. For readability, the notation ‘temporal: Observed’ is introduced in the ER diagram to denote that each relationship instance is associated with a timestamp (see literature by Fowler titled “Analysis Patterns: Reusable Object Models”).

[0059] Besides interacting with their environment, objects can also be associated with each other. The information that an object becomes the parent of another object needs to be captured. Examples of such hierarchical associations are packing (e.g. several box objects are put onto a pallet object) and product assembly (e.g. an engine block object is used to manufacture a car object). The term assembled is used to denote a hierarchical association.

[0060] Assembled. The relationship Assembled captures the fact that two objects start a parent-child-relationship at a certain time.

[0061] Disassembled. The relationship Disassembled captures the fact that two objects end a parent-child-relationship at a certain time.

[0062] It is assumed that assembly or disassembly events cannot be generated without observing the involved objects beforehand. i.e., any assembly or disassembly event is always preceded or followed by an observation event. In this case, the location where an assembly or disassembly takes place can be computed by joining the observation event with the “closest” timestamp. If assembly or disassembly does not require observation, the relationships Assembled and Disassembled can be extended to a ternary temporal relationship between objects and a location. Wang and Liu in the paper titled “Temporal Management of RFID Data” have also introduced a data model for RFID data. They allow relationships that are based on time intervals and use a relation OBJECTLOCATION to store how long an object stayed at a certain location and a relation CONTAINMENT to store how long an object was associated with another object. Some queries will benefit from such a model. However, this model introduces data dependency between organizations. Assume that manufacturer M has put a box b1 on a pallet p1, inserting a tuple (b1, p1, t1, null) into the relation Assembled. At time t2 the wholesaler W takes the box b1 of the pallet p1. Now, W has to initiate an update on M’s tuple resulting in (b1, p1, t1, t2). This requires that W has write access to M’s data which makes this model impractical for traceability applications across organizations.

[0063] Extended Traceability Model

[0064] Properties as (name, value)-pairs are introduced to allow extensions to the basic traceability model. Properties can be associated with an organization, a location, or an object (e.g. company name, location address, production date, respectively). Properties can also be associated with any of the three relationships Observed, Assembled, and Disassembled. FIG. 3 shows an expanded ER diagram. Dotted lines indicate an aggregation of a relationship.

[0065] Note that in the basic model only the appearance of an object is observed. The extended model allows additionally observed information such as temperature to be recorded as a property of the Observed relationship. Such situations are quite common. For example, real-time tracking devices fitted to cargo containers automatically collect information on each container, including physical location. parameters such as temperature and humidity and sensory readings to detect intrusion.

[0066] The model can be transformed into the following relational schema.

[0067] Organization(gid)
[0068] Location(lid, parentLid, gid)
[0069] Object(oid)
[0070] Property(name)
[0071] OrgPropertySet(oid, propertyName, value)
[0072] LocPropertySet(lid, propertyName, value)
[0073] ObjPropertySet(oid, propertyName, value)
[0074] Observed(oid, lid, ts)
[0075] Assembled(parent, child, ts)
[0076] Disassembled(parent, child, ts)
[0077] ObsPropertySet(oid, lid, ts, propertyName, value)
[0078] AsmPropertySet(oid1, oid2, ts, propertyName, value)
[0079] DsmPropertySet(oid1, oid2, ts, propertyName, value)

[0080] The attributes parent and child are oids, and the attribute ts is a timestamp.

[0081] Integration with other Data Sources

[0082] Companies will want to integrate traceability data, with existing data sources like the global data synchronization network (see the literature to Forrester Research titled “RFID: The Complete Guide”). Product information (e.g. description, brand name, color, size, and weight) can easily be represented in the current model as properties of an object. Other data sources (e.g. point of sales data) can be integrated in a similar way.

[0083] Some other data however might involve taxonomies. An example is the global trade identification number (GTIN). Each trade item has a unique GTIN. Different packaging of the same product (e.g. the retail unit itself and the case of the product) has different GTINS. FIG. 4 shows part of such a taxonomy. If the user specifies a GTIN g in his query, the answer set should take into account all GTINS within the subtree rooted by g. To support this functionality the model is extended as shown in FIG. 5. The model is extended to allow a property to be related to a taxonomy represented as a hierarchy of terms. To ensure that, a property can only be associated with the root term of a taxonomy, the relationship is constrained and allows only such instances of the entity ‘Term’ to participate in the relationship that do not have a parent term. FIG. 5 uses an OCL-like constraint specification (see Object Management Group’s OCL 2.0 Specification., Version 2.0) to impose this requirement.

[0084] The current model allows organizations to specify a query without prior knowledge about where and how data is stored within the network. However, a query specified against the global schema still needs to be executed against the local schema and data model of each organization. While a relational data model is chosen to describe the global schema an organization is free to choose a different data model (e.g. XML). Even if it chooses a relational data model, it might still choose a different schema. For example, the inventors used in experiments a different relational schema to improve the efficiency of local query processing. As only a fixed number of properties are used, a flat representation that requires less joins than the generic global representation is chosen.

[0085] Observed(oid, lid, ts, obsType, temperature)

[0086] Location (lid, locationName, locationAddress)

[0087] ObjectProperties(oid, color, maxTemp, minTemp)

[0088] To execute a global query on the local schema an organization can either provide a mechanism to rewrite global

queries to conform to their schema or provide a view on to their data that fits the global schema.

[0089] Data Ownership and Distribution

[0090] If all data is stored in a central data warehouse it is important to know which organization owns which part of the data, for example to enforce privacy policies from different organizations. If data is distributed across a network it is important to know how data is distributed to locate data sources during query processing.

[0091] An advantage of the proposed model is that, every tuple in any relation is owned by exactly one organization. It is defined which part of a relation is owned by an organization i . If not otherwise mentioned, this also corresponds to which data of a relation is located at an organization i .

[0092] It is assumed that each organization presents exactly one database to the network and describes the distribution of data with respect to gid. If each location or a subset of locations has its own database, the distribution can be analogously described with respect to lid.

[0093] All information about an organization and its locations is owned by the organization itself.

[0094] $\text{Organization}_i = \sigma_{gid=i}(\text{Organization})$

[0095] $\text{Location}_i = \sigma_{gid=i}(\text{Location})$

[0096] $\text{OrgPropertySet}_i = \sigma_{gid=i}(\text{OrgPropertySet})$

[0097] $\text{LocPropertySet}_i = \text{LocPropertySet} \bowtie_{lid} \text{Location}_i$

The relationships Observed, Assembled, and Disassembled are distributed across the network.

[0098] $\text{Observed}_i = \text{Observed} \bowtie_{lid} \text{Location}_i$

[0099] $\text{Assembled}_i = \text{Assembled} \bowtie_{closestTS} \text{Observed}_i$
where the predicate closestTS joins each assembly event with the immediately preceding observation event. Disassembled _{i} can be computed analogously to Assembled _{i} .

[0100] The organization that first introduces an object to the network (i.e., that first observes the object) owns the entry for that object in the relation Object. That is, for ownership purposes

[0101] $\text{Object}_i = \pi_{oid}(\sigma_{gid=i}(\min_{ts}(\text{Object} \bowtie_{oid} \text{Observed} \bowtie_{lid} \text{Location})))$

However, for referential integrity purposes

[0102] $\text{Object}_i = \pi_{oid}(\text{Observed})$

[0103] The relation Property is a dictionary of properties that are of global interest. Organizations can share a single instance of this relation or might replicate it.

[0104] Property entries associated with an object are owned by the organization that owns the object.

[0105] $\text{ObjPropertySet}_i = \text{ObjPropertySet} \bowtie_{lid} \text{Object}_i$

[0106] Property entries associated with observation, assembly and disassembly are owned by the organization that generated those entries. That is, such entries are distributed across the network.

[0107] $\text{ObjPropertySet}_i = \text{ObjPropertySet} \bowtie_{oid,lid,ts} \text{Observed}_i$

[0108] $\text{AsmPropertySet}_i = \text{AsmPropertySet} \bowtie_{oid,lid,ts} \text{Assembled}_i$

[0109] DsmPropertySet can be computed analogously.

[0110] Traceability Queries

[0111] How global traceability queries can be expressed in the proposed model is addressed in the following. The implications of different classes of global queries on their distributed execution are also considered. For each class, an example traceability query is given. For ease of expression

the relation ObjPropertySet is abbreviated as OPS and the relation ObsPropertySet is abbreviated as BPS.

[0112] The simplest traceability queries use only selection and projection operators.

EXAMPLE 1

List all the Temperature Readings for Objects o_1 and o_2 . The Query can be Formulated in Relational Algebra as

[0113]

$Q1: \pi_{value}(\sigma_{(oid=o_1 \vee oid=o_2) \wedge property_Name='temperature'}(BPS)).$

[0114] As described earlier, global relations can be partitioned horizontally such that each partition belongs exclusively to one organization. For queries that only use projection and selection operators it, is sufficient to execute the global query for each such partition and combine the results. It is sufficient to execute the query on partitions that have at least one tuple that matches the selection criteria. The key issue is how to detect all such partitions.

[0115] Queries with Local Join

[0116] The next class of queries additionally allows equi-joins on the attribute lid or gid. For such queries it is sufficient to execute the query at each organization and combine the results. As all data concerning a specific location or organization are owned by the same organization, no data needs to be joined between organizations. Such joins are called local joins.

EXAMPLE 2

Which Objects have Left Organization X and Y
Between Time t_1 and t_2

[0117]

$Q2: \pi_{oid}(\sigma_{(gid='X' \vee gid='Y') \wedge t1 \leq ts \leq t2}$

[0118] $(\text{Observed} \bowtie_{lid} \text{Location} \bowtie_{oid,lid,ts}$

[0119] $(\sigma_{propertyName='obsType' \wedge value='shipped'}(BPS))))$

[0120] Queries with Global Join

[0121] For all other join conditions, data from different organizations may have to be joined.

EXAMPLE 3

Which Objects Have Exceeded Their Maximum
Allowed Storage Temperature?

[0122]

$Q3: \pi_{oid}(\sigma_{Ops.value < bps.value}$

[0123] $((\sigma_{propertyName='maxTemp'}(OPS)) \bowtie_{oid}$

[0124] $(\sigma_{propertyName='temperature'}(BPS))))$

[0125] In this case, the information about the maximum allowed storage temperature is stored at the organization that created the object. It needs to be joined with observed temperature readings which are spread across all organizations on the object's path.

[0126] Queries with Aggregation

[0127] Most queries that contain aggregation can simply be executed at each organization and the overall aggregation can be computed based on the individual results. This applies for example to max, min, count, and sum. In some cases the global query might have to be extended before executing it at each location. For example, to compute an overall average, an

additional count has to be computed at each location. Details on evaluating aggregate functions in this manner can be found in the paper to Gray et al. titled “Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals”.

EXAMPLE 4

What is the Maximum Observed Temperature for Object o?

[0128]

$Q4: \max_{value}(\sigma_{property/Name='temperature' \wedge oid=o}(BPS))$

[0129] Queries with Recursion

[0130] To express traceability queries that involve assembly and disassembly of objects recursion is needed.

EXAMPLE 5

What are All the Subcomponents of Object o?

[0131] To express this query, the transitive closure of the relation Assembled is needed.

[0132] An α -extended relational algebra (see paper to Agarwal et al. titled “An Extension of Relational Algebra to Express a Class of Recursive Queries”) is used for this purpose.

$Q5: \pi_{child}(\sigma_{parent=o}(\alpha(\pi_{parent,child}(Assembled))))$

[0133] $(\alpha(\pi_{parent,child}(Assembled)))$ computes the transitive closure of Assembled using the attributes parent and child.

[0134] Recursion might also be needed when following the path of an object. Consider for example a bottle that is packed and unpacked as illustrated in FIG. 6.

EXAMPLE 6

Return all Locations the Bottle has Passed

[0135] If the bottle passes a location while it is packed into another object, it might not be observed at that location. Reader technology might only allow to observe the outermost object, e.g. only a pallet is observed not however the boxes contained in the pallet. To construct the bottle's complete path, observations that were recorded for objects which contained the bottle at the time the observation need to be accounted for.

[0136] The relation α Assembled is introduced which computes the transitive closure of Assembled. The μ -expression chooses for each tuple t the maximal timestamp of all tuples that participated in the generation of t.

α Assembled= $\alpha_{\mu,ts=\max ts(\Delta)}(\pi_{parent,child}(Assembled))$

[0137] α Disassembled is defined analogously to α Assembled, except that this time, the minimal timestamp is selected.

[0138] A relation α Contained(parent, child, lower, upper) is built that stores the time interval in which child was contained in parent by joining α Assembled and α Disassembled:

α Contained= $\pi_{A,parent,A,child,A,ts as lower,D,ts as upper}$

[0139] $(\alpha$ Assembled $\bowtie_{parent,child}$ α Disassembled D)

[0140] Based on this new relation, the example query is formulated that returns all locations the bottle passed through.

$Q6: \pi_{loc}(\sigma_{oid=o}(\text{Observed})) \cup$

[0141] $\pi_{loc}(\text{Observed} \bowtie_{oid=parent \wedge lower<timestamp<upper} (\sigma_{child=o}(\alpha$ Contained)))

[0142] Table 1, shown below, summarizes the different query classes and example queries.

TABLE 1

Example Query	Types of traceability queries.				
	w/o join	with local join	with global join	with aggregation	with recursion
Q1	X				
Q2		X			
Q3			X		
Q4	X			X	
Q5	X				X
Q6			X	X	X

[0143] System Design for Traceability Queries

[0144] Process and Forward Architecture

[0145] An architecture is introduced that combines all query capabilities in a query engine (implemented, for example, using middleware) that supports information sharing across multiple organizations. Given a traceability query, the query engine does query rewriting based on locally available information, locates remote data sources, forwards the query, and combines local and remote results. An approach is shown where the query engine is completely independent from any central services to locate the necessary data sources. However, if such services are available the query engine is able to exploit them.

[0146] The overall architecture consists of a data network and each node in the network has two components: an RFID data repository and a query processing component that is capable of retrieving data from the local repository and forward queries to other nodes in the network. A naming service exists in the network that assigns a unique address to each node for communication purposes (similar to the DNS in the Internet).

[0147] Query Processing

[0148] Each node processes a query to the extent it is able to, and forwards the query to other nodes in order to process it completely. The number of nodes accessed in order to completely process a query is kept minimal. In order to maintain the information which path an object took, each organization is required to store two additional properties for each shipping or receiving event, namely sentTo and receivedFrom (see FIG. 7 for an illustration). This information can be gained by joining traceability data with other enterprise data such as billing or accounting data. If an organization is not able to provide precise data, the number of nodes in the network that have to be visited increases. In the worst case, the network may have to be flooded.

[0149] The pseudo code provided below outlines some of the steps involved in an exemplary embodiment.

processAndForward(q)	
1	//analyze and rewrite query appropriately
2	{q _{remote} , q _{local} , q _{post} } := analyzeAndRewrite(q)
3.	//obtain results from appropriate remote nodes
4	if (q _{remote} != null)
5	r := forwardAndCombineResults(q _{remote})

-continued

processAndForward(q)	
6	//combine local results with remote results
7	if (q _{local} != null)
8	r = r ∪ executeQuery(q _{local} , localData)
9	//postprocess combined results
10	if (q _{post} != null)
11	r = executeQuery(q _{post} , r)
12	return r

[0150] The above described algorithm runs within the query engine at each organization. An incoming query is first analyzed and, depending on the outcome, rewritten. The details for the procedure `analyzeAndRewrite` are discussed below. The rewritten query is then forward to other organizations in the network and the results are united. How the relevant organizations are detected is now addressed. The algorithm for the procedure `forwardAndCombineResults` is shown above. The global query is also translated to a query that can be executed on the organization's data. Policy enforcement (see paper to LeFevre et al. titled "Limiting Disclosure of Hippocratic Databases") can be used to limit data disclosure. The local result is added to the results retrieved from the network. If necessary, postprocessing is done on the results. The details of query execution are illustrated using the example queries given previously.

[0151] Query Forwarding. As discussed in Section 5 all that needs to be done for queries Q1 and Q2 is to execute them at all relevant organizations and combine the results. Relevant nodes can be detected by analyzing all selection predicates in a query. Q2's selection predicate specifies two relevant organizations X and Y the query needs to be forwarded to.

[0152] Q1 does not specify any location or organization. However as the query specifies two objects o₁ and o₂ relevant organizations need to be on o₁'s or o₂'s path. If o₁ and o₂ are known (i.e., at least one observation event exists locally) they can be forwarded based on information specified by the properties `sentTo` and `receivedFrom`. Otherwise the query needs to be flooded to the network. If o₁ and o₂ were sent by different organizations two queries restricted to o₁ or o₂ respectively are sent out to the organizations that sent o₁ and o₂. The pseudo code of procedure `forwardAndCombineResults` is given below. Note that, although it is not shown in the pseudo code, the current implementation will group objects and send out only one query for each relevant organization.

forwardAndCombineResults(q)

1	//collect information for query routing
2	orgsRelevant := extractOrganizations(q)
3	objsRelevant := extractObjects(q)
4	//forward query to all relevant organizations
5	for each org in orgsRelevant
6	r = r ∪ forward(org, q)
7	for each obj in objsRelevant
8	q' = restrictTo(q, obj)
9	if (obj is known)
10	if (fromOrg(obj) != null)
11	r = r ∪ forward(fromOrg(obj), q')
12	if (toOrg(obj) != null)
13	r = r ∪ forward(toOrg(obj), q')
14	else

-continued

forwardAndCombineResults(q)	
15	r = r ∪ flood(q')
16	return r

[0153] Query Rewriting: Non-Local Join. Q3 cannot be executed as-is at every organization as data needs to be joined across organizations. A solution is to split the query such that each resulting query has only local joins. A postprocessing query is generated that operates on the results retrieved for the individual queries and produces the final result. Q3 is split into

q1 = σ_{property.Name = maxTemp}(OPS) and
q2 = σ_{property.Name = temperature}(BPS).

[0154] A query $q_{final} = \pi_{oid}(\sigma_{rq1.value < rq2.value}(rq1 \bowtie_{oid} rq2))$ is built and executed on the result rq1 and rq2. The pseudo code for handling non-local joins is given below in `analyzeAndRewrite`, lines 2-9.

[0155] Query Rewriting: Aggregate Functions. For the maximum function in Q4 nothing needs to be changed for the remote queries. However, a postprocessing query needs to be created that computes the overall maximum based on the individual maximum values returned. The pseudo code shown below, in lines 14-22, handles aggregation. Distributive aggregate functions like summation, algebraic aggregate functions like standard deviation and holistic aggregate functions like median have to be handled appropriately. Only the code for executing the algebraic aggregate function average is shown as an example.

[0156] Query Rewriting: Recursion. Recursion is handled by expanding the recursion as far as possible locally and adding local recursion results to the queries that are forwarded. The data shown in FIG. 7 produced by three organizations, a cylinder manufacturer CYM, an engine manufacturer ENM, and a car manufacturer CAM is used to illustrate this. An application at CAM is interested in all parts contained in object car1 (Q5). The query engine at CAM determines that the query has a recursive component $p = \alpha.Assembled$. Running Q5 on the locally available table $\alpha.Assembled$ will return $r_{temp} = engine1$. Based on this information the original query is rewritten to a query $\pi_{child}(\sigma_{parent = car1} \vee_{parent = engine1}(\alpha.Assembled))$ that will be forwarded to other nodes.

[0157] The query is forwarded to ENM. At ENM the same process takes place and ENM detects that `engine1` contains `cylinder1` and `cylinder2`. The query is forwarded to CYM where forwarding terminates as the recursion is no longer expanded.

[0158] The pseudo code for handling recursion is given below, lines 10-13:

analyzeAndRewrite(q)

1	q _{remote} := q
2	//handle non-local joins
3	qSet := splitIntoLocalJoinQueries(q)
4	rSet := ∅
5	q _{final} := buildFinalQuery(q, qSet)
6	for each q _{split} in qSet
7	rSet = rSet ∪ processAndForward(q _{split})
8	return executeQuery(q _{final} , rSet)
9	//algorithm terminates
10	//handle recursion
11	for each recursive component p in q

-continued

analyzeAndRewrite(q)	
12	$r_{temp} := \text{executeQuery}(p, db_{local})$
13	$\text{addResultsToRemoteQuery}(q_{remote}, p, r_{temp})$
14	//handle aggregate functions
15	for each aggregate function f in q
16	
17	if (f is $\text{avg}(X)$)
18	//need count to compute overall average
19	$\text{addQueryFragment}(q_{remote}, \text{"count(X)"})$
20	//determine postprocessing
21	$\text{addToPostprocess}(q_{post}, f)$
22	//process other aggregate functions
23	//build local query
24	$q_{local} := \text{mapToLocalSchema}(q_{remote})$
25	$q_{local} = \text{enforcePolicy}(q_{local})$
26	return $\{q_{remote}, q_{local}, q_{post}\}$

[0159] The present invention, therefore, provides for a generic query engine for sharing RFID data across independent organizations. A new paradigm is developed for processing queries across a network of data sources. A query is first processed by a data source and based on locally available information rewritten and forwarded to other data sources. Some of the advantages of this process-and-forward solution over previous approaches are outlined below.

[0160] The introduced query engine supports all traceability queries. This enables rapid development of applications on top of it as no query functionalities need to be implemented within the application.

[0161] As the amount of RFID data coming online increases, the total data transferred may put serious constraints on approaches that rely on data transfer to a central repository or a central directory service. The presented query engine does not require any data to be transferred except for queries and query results.

[0162] There is a single point of failure in the central warehouse as well as the EPCglobal approach. If the warehouse or the core services are not available, no queries can be executed. If an organization's query engine is not available in the process-and-forward approach, it only affects queries that need the offline node. If a node on the path of an entity fails, the process-and-forward approach can still process the query up to the failing node.

[0163] The process-and-forward approach leaves all control over the data to the individual organization. EPCglobal and central warehouse require organizations to surrender some or all control over their data and trust a third-party to enforce all their confidentiality requirements.

[0164] Although current RFID technology has trouble reading tags that are obscured by a container, previous approaches did not talk about how objects that are packed and unpacked can be traced across organizations. The present invention's model together with the presented algorithms provides a solution to all containment related issues. The presented solutions allow for the first time to trace objects that get packed and unpacked.

[0165] Additionally, the present invention provides for an article of manufacture comprising computer readable program code embodied therein implementing a middleware to query RFID data across a network of independently operated nodes sharing RFID data. Furthermore, the present invention includes a computer program code-based product, which is a storage medium having program code stored therein which

can be used to instruct a computer to perform any of the methods associated with the present invention. The computer storage medium includes any of, but is not limited to, the following: CD-ROM, DVD, magnetic tape, optical disc, hard drive, floppy disk, ferroelectric memory, flash memory, ferromagnetic memory, optical storage, charge coupled devices, magnetic or optical cards, smart cards, EEPROM, EPROM, RAM, ROM, DRAM, SRAM, SDRAM, or any other appropriate static or dynamic memory or data storage devices.

[0166] Implemented in the middleware are: (a) computer readable program code to aid in receiving a query; (b) computer readable program code to aid in retrieving data related to the query from a local RFID repository; (c) computer readable program code to analyze, execute and, optionally, rewrite the query based on retrieved data; and (d) computer readable program code to aid in forwarding the rewritten query to another of the independently operated node, wherein the query is efficiently processed by combining results of query executions across a plurality of independent nodes.

CONCLUSION

[0167] A system and method has been shown in the above embodiments for the effective implementation of a middleware for query processing across a network of RFID databases. While various preferred embodiments have been shown and described, it will be understood that there is no intent to limit the invention by such disclosure, but rather, it is intended to cover all modifications and alternate constructions falling within the spirit and scope of the invention, as defined in the appended claims.

1. A network of independently operated nodes sharing RFID data, each node comprising:

(a) a local RFID repository; and

(b) a query engine, said query engine

receiving a query,

retrieving data related to said query from said local RFID repository,

analyzing, executing, and optionally rewriting said query based on retrieved data, and

forwarding said rewritten query to another of said independently operated node,

wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

2. A network of independently operated nodes sharing RFID data, as per claim 1, wherein said query engine is implemented as middleware.

3. A network of independently operated nodes sharing RFID data, as per claim 1, wherein said local RFID repository stores at least entity type and event type data.

4. A network of independently operated nodes sharing RFID data, as per claim 3, wherein said entity type is any of the following: a uniquely identifiable object, a uniquely identifiable location, or a uniquely identifiable organization.

5. A network of independently operated nodes sharing RFID data, as per claim 4, wherein said event type is any of the following:

observed event type identifying a ternary relationship between an object, location and a timestamp,

assembled event type identifying a start of a parent-child relationship between two objects, or

disassembled event type identifying when a parent-child relationship ends between two objects.

6. A network of independently operated nodes sharing RFID data, as per claim 3, wherein said event type or entity type further identifies at least one property as a name, value pair.

7. A network of independently operated nodes sharing RFID data, as per claim 6, wherein said property is related to a taxonomy comprising a hierarchy of terms.

8. A network of independently operated nodes sharing RFID data, as per claim 1, wherein each node independently enforces a restriction policy limiting data disclosure.

9. A network of independently operated nodes sharing RFID data, as per claim 1, wherein said combined results are used for any of the following: pedigree generation, targeted recall of a product, detection of unusual patterns, or counterfeit detection.

10. A method implemented across a network of independently operated nodes sharing RFID data, said method implemented as middleware in a node among said network of independently operated nodes comprising the steps of:

- (a) receiving a query;
- (b) retrieving data related to said query from a local RFID repository;
- (c) analyzing, executing and, optionally, rewriting said query based on retrieved data; and
- (d) forwarding said rewritten query to another of said independently operated node,

wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

11. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 10, wherein said local RFID repository stores at least an entity type and an event type data.

12. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 11, wherein said entity type is any of the following: a uniquely identifiable object, a uniquely identifiable location, or a uniquely identifiable organization.

13. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 12, wherein said event type is any of the following:

- observed event type identifying a ternary relationship between an object, location and a timestamp,
- assembled event type identifying a start of a parent-child relationship between two objects, or
- disassembled event type identifying when a parent-child relationship ends between two objects.

14. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 11, wherein said entity type or said event type further identifies at least one property as a name, value pair.

15. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 14, wherein said property is related to a taxonomy comprising a hierarchy of terms.

16. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 10, wherein each node independently enforces a restriction policy limiting data disclosure.

17. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 10, wherein said combined results are used for any of the following: pedigree generation, targeted recall of a product, detection of unusual patterns, or counterfeit detection.

18. An article of manufacture comprising a computer user medium having computer readable program code embodied therein implementing a middleware to query RFID data across a network of independently operated nodes sharing RFID data, wherein said middleware comprises:

- (a) computer readable program code aiding in receiving a query;
- (b) computer readable program code aiding in retrieving data related to said query from a local RFID repository;
- (c) computer readable program code analyzing, executing and, optionally, rewriting said query based on retrieved data; and
- (d) computer readable program code aiding in forwarding said rewritten query to another of said independently operated node,

wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

19. A method implemented across a network of independently operated nodes sharing RFID data comprising the steps of:

- (a) receiving a query;
- (b) analyzing and rewriting said query;
- (c) retrieving data related to said rewritten query from a local RFID repository and evaluating local results;
- (d) identifying additional nodes for forwarding said rewritten query;
- (e) forwarding said rewritten query to said identified nodes, where said rewritten query is evaluated using data in corresponding local RFID repositories to produce remote results;
- (f) obtaining results of evaluated query from said identified nodes; and
- (g) combining said evaluated local results with said remote results,

wherein said query is efficiently processed by combining results of query executions across a plurality of independent nodes.

20. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 19, wherein said local RFID repository stores at least an entity type and an event type data.

21. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 20, wherein said entity type is any of the following: a uniquely identifiable object, a uniquely identifiable location, or a uniquely identifiable organization.

22. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 21, wherein said event type is any of the following:

- observed event type identifying a ternary relationship between an object, location and a timestamp,
- assembled event type identifying a start of a parent-child relationship between two objects, or
- disassembled event type identifying when a parent-child relationship ends between two objects.

23. A method implemented across a network of independently operated nodes sharing RFID data, as per claim 20,

wherein said entity type or said event type further identifies at least one property as a name, value pair.

24. A method implemented across a network of independently operated nodes sharing RFID data, as per claim **23**, wherein said property is related to a taxonomy comprising a hierarchy of terms.

25. A method implemented across a network of independently operated nodes sharing RFID data, as per claim **19**, wherein each node independently enforces a restriction policy limiting data disclosure.

* * * * *