# X-CleLo: Intelligent Deterministic RFID Data Transformer

Peter Darcy, Bela Stantic and Abdul Sattar

Institute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia

**Abstract.** Recently, passive Radio Frequency Identification (RFID) systems have received an exponential amount of attention as researchers have worked tirelessly to implement a stable and reliable system. Unfortunately, despite vast improvements in the quality of RFID technology, a significant amount of erroneous data is still captured in the system. Currently, the problems associated with RFID have been addressed by cleaning algorithms to enhance the data quality. In this paper, we present X-CleLo, a means to intelligently clean and enhance the dirty data using Clausal Defeasible Logic. The extensive experimental study we have conducted has shown that the X-CleLo method has several advantages against a currently utilised cleaning technique and achieves a higher cleaning rate.

## 1 Introduction

RFID is technology which uses radio communication between tags and readers to automatically identify the locations of items. The architecture itself has the potential to simplify the processes currently involved in capturing data and, eventually, replace the barcode system employed in various distribution retailers. Despite significant advances in technology, there are still problems surrounding the implementation of the system. These problems include: the data being at low level; the large amount of incorrect readings being recorded; the exponential volume of data continually stored; and the complex temporal and spatial aspects of the data.

Current methods employed to enhance RFID data include using deterministic rules to improve the integrity of the observations. While this method does provide a means of enhancing the RFID data's accuracy, there are fatal flaws within the said method that undermine the integrity of the resulting data sets. These flaws result from the rule based approach lacking a higher level of intelligence needed when faced with an ambiguous situation which can result in no detection of various anomalies or the creation of artificial anomalies.

A common solution put forward within the data warehouse community is that dirty data is extracted from the data warehouse, transformed to comply with a set of rules determined by the user, and loaded back into the data warehouse. The rules employed in the transformation stage may, unfortunately, be undermined due to an ambiguous situation. A method to cope with ambiguous situations is to use *Clausal Defeasible Logic* to find the correct solution when given several facts. Clausal Defeasible Logic is a theory of *Non-Monotonic Reasoning* that has been devised to seek out the correct solution when having passed several different inputs using various levels of ambiguity.

Darcy P., Stantic B. and Sattar A.
X-CleLo: Intelligent Deterministic RFID Data Transformer.
DOI: 10.5220/0003015700030012
In *Proceedings of the 4th International Workshop on RFID Technology - Concepts, Applications, Challenges (ICEIS 2010)*, page
ISBN: 978-989-8425-11-9

In this work, we propose X-CleLo (eXtra-CLEan-LOad), an intelligent and deterministic method to enhance the integrity of the RFID observations. X-CleLo has been designed to clean the records intelligently in order to transform the said data into highly accurate observations worthy of being used in a commercial environment. Through an experimental study on a simulated RFID hospital scenario, we have demonstrated how X-CleLo houses distinct advantages that allows it to perform better than the currently employed rule-based approach.

The remainder of this paper is organised as follows: Section 2 will discuss the fundamental concepts needed to understand the mechanics of X-CleLo; Section 3 reviews some of the current state-of-the-art related work used to achieve clean RFID data; Section 4 describes the architecture of X-CleLo and the scenario considered in experiment; Section 5 discusses the experimental evaluation we have run on X-CleLo; In Section 6, we analyse the experiments performed using X-CleLo and in Section 7, we conclude this study as well as provide some directions for future work.

## 2 Background

Acceptance and implementation of RFID systems is still hindered due to the complexities associated with integrating the technology into the environment. Due to the nature of the capturing process, certain anomalies such as dirty data prevent attempts to create a flawless RFID system which would usually be handled by data correction methodologies.

### 2.1 RFID

Radio Frequency Identification (RFID) is technology created to identify automatically a group of individual items. The RFID architecture is comprised of three main components: the tag, the reader, and the middleware used to interpret the tag data found within the proximity of readers. RFID systems have been integrated into various applications, for example, Supply-Chain situations [1] and Hospital Patient Monitoring [2]. The advent of RFID technology has also ushered in a promise of simplification of time-costly tasks. However, wide-scale deployment appears to have been stalled due to the problems in the implementation phase.

These problems are most prevalent within RFID applications which use the passive tags (tags that do not contain power source). Passive tags, however, have several advantages such as an affordable price, ease of integration into applications and, theoretically, they will last forever. Four aspects of RFID systems have been identified as problematic: low level representations of the observations, error-prone data, high information collection volumes and complex spatial and temporal aspects of the data, [3], [1] and [4].

Data which is captured by the readers has been found to be erroneous in providing *False Negative Readings* (readings which should have been recorded but were not), *False Positive Readings* (readings which were not supposed to be recorded, but they were) and *Duplicate Readings* (readings in which two identical records exist where only one should). It has been estimated that only 60%-70% of tags are read when an RFID scan is performed [5], [6].

### 2.2 Clausal Defeasible Logic

Clausal Defeasible Logic is a type of Non-Monotonic Reasoning that has been developed to obtain a conclusion after given certain observations and rules [7]. For example, in Figure 1, it has been shown that mammals do not usually lay eggs. There is an exception to this rule however, as certain creatures, such as the platypus, do lay eggs and are still mammals [8]. It is important to note the relation represented as a semi circle giving priority to the most clockwise entity, in this case the platypus rule, in this example. This relation is known as a "Priority Relation"; in a typical scenario, priority is given to the rule which is more specific and will always reach the conclusion before other rules.

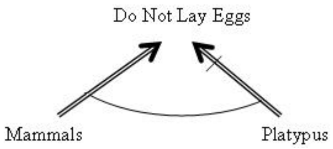

**Fig. 1.** The Clausal Defeasible Logic map of Platypus and Mammals laying eggs.

From the rules shown in Figure 1, the Clausal Defeasible Reasoning Engine can ascertain that platypus are mammals and, therefore, must not lay eggs. However, since the second rule which states that for the platypus, the negative of "don't lay eggs" is defeasible, it may be stated that the platypus does lay eggs. Although it is true that one conclusion must be drawn for any given situation, Clausal Defeasible Logic, has several levels of confidence represented in formulae that can be used to obtain a different correct answer dependent on the amount of ambiguity allowed [8]. These formulae include:

- $\mu$**:** This formula uses only certain information to obtain its conclusion.
- $\pi$**:** This formula allows conclusions in which ambiguity is propagated.
- $\beta$**:** This formula does not allow any ambiguity to be used to obtain its conclusion.

## 3 Related Work

There have been many different methods proposed in literature pertaining to how to deal with RFID's problems. General methodologies used to correct dirty data include conditional functional dependencies [9] and data repairing via updates [10]. Although these methodologies clean the majority of anomalies in the average data set, they would struggle to clean the unique characteristics of dirty RFID observations. Some cleaning methods which have been specifically tailored to correct anomalous data in RFID applications include filtration algorithms [6], probabilistic approximations [11], [12], a sampling-based approach [13] and a deferred cleansing rule-based method [14]. We have chosen to investigate the "Deferred Cleansing Method" as it cleanses data at a deferred stage of the capturing cycle.

### 3.1 Deferred Cleansing

The "Deferred Cleansing Method" is a method which uses a rewrite engine to correct anomalies within the database after the capture process [14]. This is done by the system referring to several rules defined by the user which can be applied in a specific order in an attempt to correct the data. An example rule is that, if there are two readings of the same tag within one minute at the same location, the system will treat the second tag as a duplicate entry and promptly proceed to eliminate it.

We have found a fundamental flaw, however, in that the rule cleaning process does not take into consideration ambiguous situations. In the event that one or more rule(s) conflict, the order that the rules have been implemented dictates how the data set will be cleaned. An example of this flaw would be if there is a rule which states that a second tag should be deleted. If it is detected at two different reader locations far apart, then the rule will move on to delete one. However, in reality, the situation could arise that it is the second case which reflects the true location of the tag and that the first itself is the wrong data.

## 4 Extra-Clean-Load (X-CleLo)

In this section we present the core of our method for enhancing RFID data, eXtra-CLEan-LOad (X-CleLo) which was designed specifically to emphasise the cleaning of data intelligently. The particular information we will focus on providing in this section includes a detailed analysis of X-CleLo's architecture, the Logical Engine structures and assumptions we have made while performing experimentation. With regards to the actual storing of the RFID readings, we have closely followed database structure introduced by [15] called the Data Model for RFID Applications (DMRA).

### 4.1 X-CleLo Structure

As seen in Figure 2, X-CleLo is divided up into three main sections - the *Extraction* processes, the *Transformation* processes and the *Loading* process. To accomplish this, it requires three sets of input information to properly function. These inputs include the readings, the referential data and the user's input. The data is the recorded readings of RFID tags. The RFID data consists of the unique tag identifier, the timestamp the tag was recorded at and the reader identifier.

The referential data set contains all related data which will be used later to transform the low level readings into high level events. Additionally, this data will include a MapData table possessing a list of readers which are geographically within proximity. These information sets may include the location of the reader, the items which the tag is attached to etc. The User is the person who inputs various event rules into the system through use of the Non-Monotonic Reasoning Engine.

After the input information has been received, X-CleLo runs the "Extract" phase which consists of two information-processing function, the *Data Extractor* and the *Rule Extractor*. The Data Extractor is the process of accepting the "RFID Readings" and the "RFID Referential Data" and mining other tables for information related to the data.

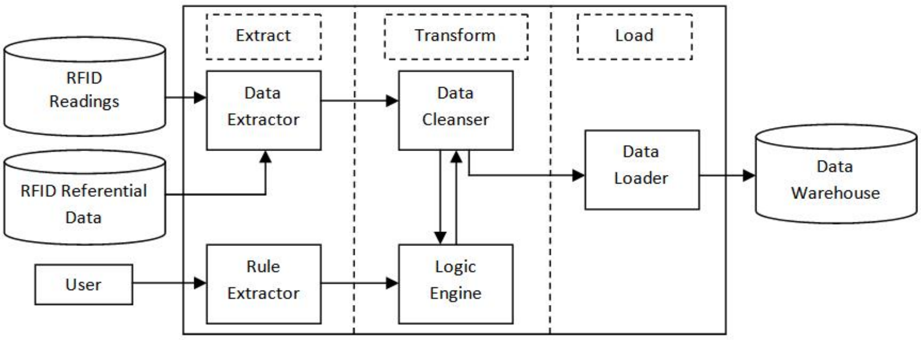

**Fig. 2.** A high level diagram to represent the processes and data flow of X-CleLo.

The Rule extractor receives the logic engine setup devised by the "User" and forwards it onto the Non-Monotonic Reasoning Engine.

The next phase which occurs in X-CleLo is the "Transformation" phase in which the information is then turned into higher, more meaningful data. The processes used to accomplish this include the *Data Cleanser* and the *Non-Monotonic Reasoning Engine*. The Data cleansing process consists of an algorithm which prepares the data for the next phase. It uses the Non-Monotonic Engine to find the correct course of action of cleaning when ambiguous situations arise. It employs a set of logic in the form of Clausal Defeasible Logic.

The specific data to be cleaned is taken from the "Data Extractor" and, specifically, is the "RFID Readings." The Non-Monotonic Reasoning Engine houses the Clausal Defeasible Logic program discussed in Section 2.2 which is used to determine the correct course of action when given information. The rules are directly stated by the user. The corrected data is then passed to the *Data Loader* process in the "Load" phase which will then load the corrected information into the database.

### 4.2 Logic Engine Structure

X-CleLo relies on Clausal Defeasible Logic Engines to determine the correct outcome when given observational inputs. We have set up two logic engines to deal with missing and wrong values from the RFID readings.

The first logic engine depicted in Figure 3 is the *Wrong Value Logic Map*. It is devised to detect and remove any readings that have been found to be recorded incorrectly when analysing other observations. The engine will then check if there are no readers that can fill in the extra value but still be geographically correct according to the MapData table. If it cannot find any reader to fill the gap left by the extra value, the program will promptly delete it. Before it deletes the said value, it will check to see if there are any combinations outside the temporally closest readings.

The second engine shown in Figure 4 is the *Missing Value Logic Map* which has been designed to be used when cleaning the RFID readings and makes use of the MapData table. It operates by examining if the missed value (detected due to the assumption that readers scan every minute) is between two values that, if, when an additional read-

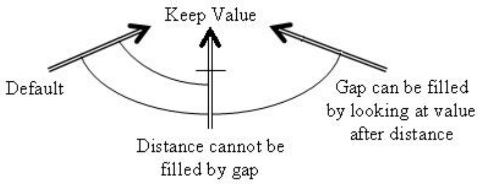
Keep Value
Default
Gap can be filled by looking at value after distance
Distance cannot be filled by gap

**Fig. 3.** The Logic Map created with the intent of detecting and deleting wrong values.

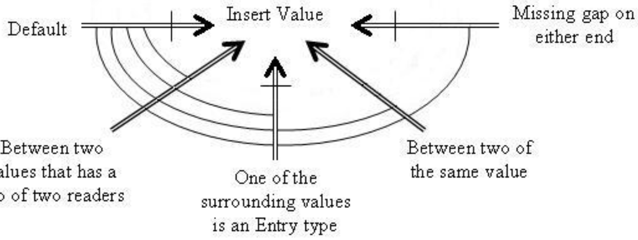
Default
Insert Value
Missing gap on either end
Between two values that has a gap of two readers
One of the surrounding values is an Entry type
Between two of the same value

**Fig. 4.** The Logic Map created with the intent of detecting and inserting missing values.

ing is inserted, would be next to each other geographically according to the MapData table.

If the case is that the inserted value would geographically make the observation sound, it will conclude that it needs to insert the value. It then searches to see if the readers around the missed value are temporally next to an entry/exit reader and, if so, it will conclude that the value will not be inserted. The logic engine then checks if the value is between two of the same readers and, if so, concludes to insert the value, and finally, if the value is temporally between two other missed values, it will conclude that the value will not be inserted.

## 5 Experimental Evaluation

As mentioned earlier, X-CleLo makes use of the DMRA to store all relevant RFID data. The tables found in DMRA and the procedural language PL/SQL were implemented on Oracle 10g RDBMS. The Clausal Defeasible Non-Monotonic Reasoning Engine is written in the *C* language, implemented on Cygwin Version 1.5.24-2. In the absence of real data, we generated data to simulate a hospital scenario based on real world applications [2]. The type of simulated data includes a unique tag id, timestamp as to when the reading occurred, and the identifier of the reader. To extensively experiment X-CleLo, we considered four data sets changing the number of deletions in the observation table. We have devised an algorithm to examine each record and to delete randomly the record based on a deletion chance percentage threshold system to simulate a real RFID situation with missing data. Due to the previously mentioned statement that only 60%-70% RFID tags are captured, the four data sets have been given a different deletion percentage, namely 20%, 30%, 40% and 50%.

## 6 Results and Analysis

For the experimental study of the effectiveness of X-CleLo, we ran two experiments to compare X-CleLo to a naive rule based method. The naive rule based algorithm uses rules in the order used by X-CleLo. However, it lacks defeasible logic to ascertain high level intelligence when cleaning. Through this experiment, we intend to identify X-CleLo's intelligence cleaning success while compared to the rule based algorithms such as the rules used in the Deferred Cleansing method.

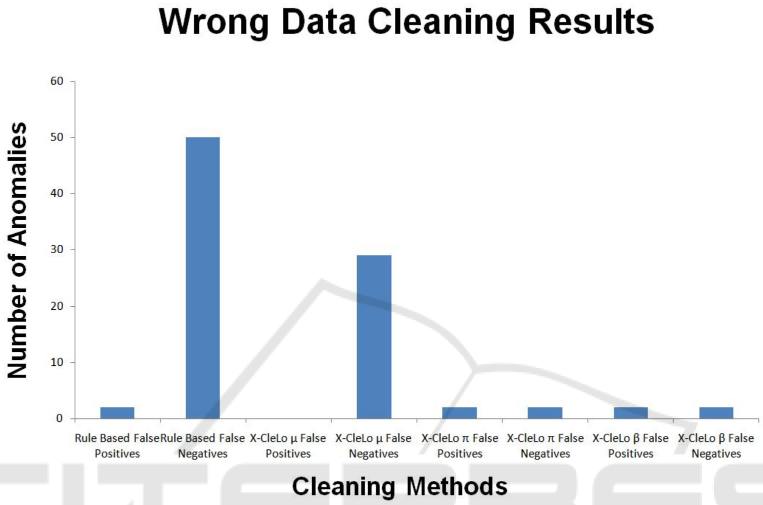

**Fig. 5.** The results of cleaning the Observation table with wrong data present. Bars are divided up into the amount of false positive and false negatives for each of the *Rule Based* and the *Clausal Defeasible Formulae* ($\mu$ $\pi$ and $\beta$).

### 6.1 Wrong Data Results

The first experiment tests X-CleLo against the *Rule Based Algorithm* when faced with the situation of having to clean false positive errors. In this experiment, a number of randomly artificially produced false positive readings are introduced into the `Observation` table. X-CleLo and the *Rule Based Algorithm* are then subjected to the dirty data set to attempt to clean it and restore its integrity as best they can by deleting the false positive readings. In this experimentation, we term false positives as any readings which were not present within the original data set, and false negatives as any data which has been deleted that was present within the original data set.

The graph of the results may be seen in Figure 5 in which the amount of correct deletions and the amount of false positives still present are recorded for each algorithm. It is important to note that the least performing algorithms in both false positives and negatives were the $\mu$ Clausal Defeasible formula and the Rule Based which had the highest false positive anomalies and false negative anomalies respectively. Both the $\pi$ and $\beta$ Clausal Defeasible formulae proved to have the smallest amount of false positive and false negative anomalies obtaining the lowest false negatives and positives each.

### 6.2 Missing Data Results

The second experiment was to evaluate the performance of X-CleLo against the *Rule Based Algorithm* when faced with the situation of having to clean missed readings. In this experiment, four data sets were generated each with the complete scenario but having different percentages of original values still present. The missed values are simulated in the data sets by choosing a random number between 1% and 100%. If the random number is below the specified threshold percentage unique to each data set, the reading will be deleted. The thresholds used are based upon the statement that only 60% to 70% of RFID readings are recorded. We have set the threshold to be 50%, 60%, 70% and 80% which results in approximately 50%, 40%, 30% and 20% of the data being deleted respectively. The *X-CleLo* and *Rule Based* Algorithms are then employed to seek out missed readings and infer them using their respected methods. From analysing the accuracy of the data sets after performing the cleaning in the graph in Figure 6, we found that the rule based algorithm and X-CleLo $\mu$ performed the least successfully while the X-CleLo $\pi$ and X-CleLo $\beta$ formulae provided the most thorough clean. As a general observation, it appears that in every test the highest accuracy is achieved with lower missing values.

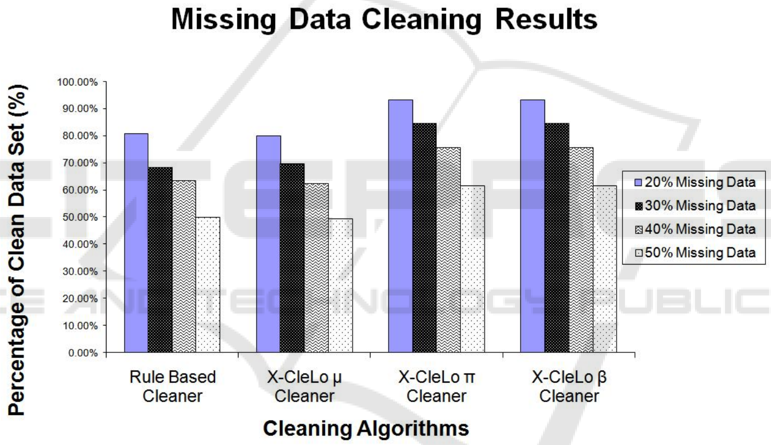

**Fig. 6.** The results of cleaning the Observation table with missing data present. Results are divided into the performance of the Rule Based and Clausal Defeasible Formulae ($\mu$ $\pi$ and $\beta$) when attempting to clean a database with 50%, 40%, 30% and 20% of the data missing.

### 6.3 Analysis

From the above results, it may be detected that there is a common element in all the tests performed, X-CleLo's $\pi$ and $\beta$ formulae provided the greatest accuracy followed by either the X-CleLo's $\mu$ formula or the algorithm used for comparison (rule based or probabilistic). The reason why the $\pi$ and $\beta$ formulae have achieved a higher integrity level than that of the $\mu$ formula may be explained in that the $\mu$ formula will only allow strict rules to direct the answer to a conclusion. The logic engines used in this experiment also contributed to the $\mu$ formula not functioning. If a strict rule had been in place rather than plausible rules, the $\mu$ formula would have sought an answer.

We have found two core strengths while using X-CleLo as a cleaning method. The first is that we are able to obtain a higher level of accuracy when compared with other techniques which state-of-the-art methodologies employ. We accomplish this by providing the methodology in which rules are stated by the user and employed on the data set with high level intelligence in the form of Non-Monotonic Reasoning. The second strength is that the utilisation of a deterministic cleaning approach is a way of ensuring that the lowest amount of false positives are introduced into the data set via cleaning. This has been reflected within our first experiment in which there are a large amount of false negatives introduced when the rule based approach is used. We counter the effects of introducing false negative readings by applying Non-Monotonic Reasoning to justify when and where it is unacceptable to delete a value that appears ambiguously placed.

## 7 Conclusions

In this paper we addressed the issues related to the problems associated with errors in RFID data. Specifically, this study makes the following contributions to the field:

- We have proposed X-CleLo, a deterministic method that emphasises the cleaning of the data set intelligently by incorporating Clausal Defeasible Logic.
- We have demonstrated that X-CleLo can be effectively used to clean stored RFID data.
- We have identified that the $\pi$ and $\beta$ Clausal Defeasible Logic formulae achieved the highest accuracy when attempting to clean the erroneous data.
- In the experimental study, we compared the performance of X-CleLo to the state-of-the-art Rule-Based cleaning system. Experimental results have shown that X-CleLo obtained a higher cleaning rate.

The proposed concept is not necessarily limited to the RFID applications and may be applied as a base principle in resolving ambiguous data anomalies, specifically, large databases that have spatio-temporal recordings with false-positive and false-negative data anomalies. In terms of future work, we have identified two enhancements to improve the accuracy of X-CleLo. The first is to develop a method which will allow various rule manipulations with the logic engines in order to calibrate X-CleLo. The second is to devise increasingly complex logic maps, which will be able to correct complex ambiguous observations for scenarios outside the scope of our experimental evaluation.

## Acknowledgements

This research is partly supported by ARC (Australian Research Council) grant number DP0557303.

## References

1. Derakhshan, R., Orlowska, M. E., Li, X.: RFID Data Management: Challenges and Opportunities. In: RFID 2007. (2007) 175 – 182

2. Swedberg, C.: Hospital Uses RFID for Surgical Patients. RFID Journal (2005) Available from: http://www.rfidjournal.com/article/articleview/1714/1/1/.
3. Cocci, R., Tran, T., Diao, Y., Shenoy, P. J.: Efficient Data Interpretation and Compression over RFID Streams. In: ICDE, IEEE (2008) 1445–1447
4. Wang, F., Liu, P.: Temporal Management of RFID Data. In: VLDB. (2005) 1128–1139
5. Floerkemeier, C., Lampe, M.: Issues with RFID usage in ubiquitous computing applications. In Ferscha, A., Mattern, F., eds.: Pervasive Computing: Second International Conference, PERVASIVE 2004. Number 3001, Linz/Vienna, Austria, Springer-Verlag (2004) 188–193
6. Jeffery, S. ., Garofalakis, M. N., Franklin, M. J.: Adaptive Cleaning for RFID Data Streams. In: VLDB. (2006) 163–174
7. Billington, D.: Propositional Clausal Defeasible Logic. In: European Conference on Logics in Artificial Intelligence (JELIA). (2008) 34–47
8. Billington, D.: An Introduction to Clausal Defeasible Logic [online]. David Billington's Home Page (2007) Available from: http://www.cit.gu.edu.au/∼db/research.pdf.
9. Golab, L., Karloff, H., Korn, F., Srivastava, D., Yu, B.: On Generating Near-Optimal Tableaux for Conditional Functional Dependencies. VLDB Endow. 1 (2008) 376–390
10. Wijsen, J.: Database repairing using updates. ACM Trans. Database Syst. 30 (2005) 722–768
11. Ré, C., Letchner, J., Balazinksa, M., Suciu, D.: Event Queries on Correlated Probabilistic Streams. In: SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, New York, NY, USA, ACM (2008) 715–728
12. Tran, T., Sutton, C., Cocci, R., Nie, Y., Diao, Y., Shenoy, P.: Probabilistic Inference over RFID Streams in Mobile Environments. In: ICDE '09: Proceedings of the 2009 IEEE International Conference on Data Engineering, Washington, DC, USA, IEEE Computer Society (2009) 1096–1107
13. Xie, J., Yang, J., Chen, Y., Wang, H., Yu, P. S.: A Sampling-Based Approach to Information Recovery. In: ICDE '08: Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Washington, DC, USA, IEEE Computer Society (2008) 476–485
14. Rao, J., Doraiswamy, S., Thakkar, H., Colby, L. S.: A Deferred Cleansing Method for RFID Data Analytics. In: VLDB. (2006) 175–186
15. Wang, F., Liu, S., Liu, P.: A temporal RFID data model for querying physical objects. Pervasive and Mobile Computing In Press, Corrected Proof (2009)