



*Citation for published version:*

Woodall, P, Giannikas, V, Lu, W & McFarlane, D 2019, 'Potential Problem Data Tagging: Augmenting information systems with the capability to deal with inaccuracies', *Decision Support Systems*, vol. 121, pp. 72-83. <https://doi.org/10.1016/j.dss.2019.04.007>

*DOI:*

[10.1016/j.dss.2019.04.007](https://doi.org/10.1016/j.dss.2019.04.007)

*Publication date:*

2019

*Document Version*

Peer reviewed version

[Link to publication](#)

*Publisher Rights*

CC BY-NC-ND

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Potential Problem Data Tagging: Augmenting information systems with the capability to deal with inaccuracies

Philip Woodall\*<sup>1</sup>, Vaggelis Giannikas<sup>2</sup>, Wenrong Lu<sup>1</sup>, Duncan McFarlane<sup>1</sup>

1. Department of Engineering, 17 Charles Babbage Road, Cambridge, CB3 0FS, U.K.

2. School of Management, University of Bath, Bath, BA2 7AY, U.K.

Keywords: Data quality; information quality; accuracy; metadata; data analytics; data tags.

\*Corresponding author

**Abstract** Data quality tags are a means of informing decision makers about the quality of the data they use from information systems. Unfortunately, data quality tags have not been successfully adopted despite their potential to assist decision makers. One reason for the non-adoption is that maintaining the tags is expensive and time-consuming: having a tag that represents accuracy, for example, would be massively time-consuming to measure because it requires some physical observation of reality to check the true value. We argue that a useful surrogate tag for accuracy can be created—without having to physically measure it—by counting the number of times the data has been exposed to an event that could cause it to become inaccurate. Experimental results show that the tags can help to avoid problems caused by inaccuracies, and also to help find the inaccuracies themselves.

## 1 Introduction

Data quality tags have been proposed as a means of informing decision makers about the quality of the data they are using from information systems (Shankaranarayan et al., 2003; Wang et al., 1995)—tags are also referred to as data quality metadata (Shankaranarayanan and Even, 2006). Tags are important because without them, information systems essentially present data to users asserting that it is 100% correct, with no indication being provided about any deficiencies in the data (Kent, 1991; Motro and Rakov, 1997; Parssian et al., 2015; Yeganeh et al., 2014). If these deficiencies remain hidden, any resulting data analytics or decision-

making from the data will be flawed (Ardagna et al., 2018) leading to costly, disruptive or even disastrous situations for both business (Fisher and Kingma, 2001; Hüner et al., 2011; Redman, 1998) and academia (Marsden and Pingry, 2018).

Despite their importance, data quality tags have not been successfully adopted in information systems. One of the key barriers is that it is expensive and time-consuming to maintain such information (Chengalur-Smith et al., 1999; Fisher et al., 2003; Price and Shanks, 2011). For instance, it is notoriously difficult to measure and maintain a data tag about data accuracy (Hipp et al., 2007), which is a key data quality dimension. This is because data accuracy problems are defined as the degree to which the reported value is in conformance with the actual or true value (Ballou and Pazer, 1995; English, 1999), and so to detect the presence of an inaccuracy often requires some manual observation or survey of the real or true value (English, 2003; Heinrich and Klier, 2015). The actual value is often unknown (Joglekar et al., 2013; Shankaranarayan et al., 2003) and hence measuring inaccuracies is often time-consuming, labour intensive (Han et al., 2010), error-prone in itself (Olson, 2003), and very difficult/costly to automate (Heinrich and Klier, 2015).

In order to address this issue, we argue that a useful tag for accuracy can be created and used in information systems that does not rely on performing physical observations. By counting the number of times the data has been exposed to an event that can cause it to become inaccurate, a useful surrogate for an accuracy measurement can be obtained automatically; we refer to this as Potential Problem Data Tagging, and show how it can be used to help avoid operational disruptions caused by inaccuracies and also help to find the inaccuracies in information systems. This work builds on the concepts from currency prediction work (Heinrich and Klier, 2015; Zak and Even, 2017) that provides models to estimate how data will degrade over time, without having to rely on a baseline (e.g. manual observation) for comparison.

Potential Problem Data Tagging does not produce an bona fide accuracy measurement that can be used to assess accuracy in the sense of a normal metric because it is essentially a best guess at whether the data

is in error or not, and it could be inaccurate in itself. Hence, we also show how this surrogate measure can be used in an operational decision-making environment to help select the best course of action from a set of alternative actions; in order to improve the performance of the action (such as by avoiding operational disruptions). In summary, the main contributions of this paper are: 1. showing how to automatically populate tag values that can be used as a substitute for an inaccuracy measure, and 2. showing how these values can be used within an operational environment to select the best course of action for a decision.

Using an experimental simulation, based on a real-life decision-making task of choosing what locations to pick items from in a warehouse, we explore the cases when this type of approach can provide the most useful indicators to help decision makers avoid problems with inaccurate data. Although the approach leads to recording many false positives, the experimental results show that the tags are usable and perform the best for high error rates and for problems with larger degrees of freedom with respect to the decision alternatives.

### 1.1 Motivating example: Inaccuracies in warehouse decision-making

One example of a real decision-making case which can benefit from accuracy data tags is choosing what location to pick items from within a warehouse. In this case, a warehouse management system (WMS) records the locations, types and quantity of items in the warehouse. The decision uses the type and quantity criteria to determine which location to pick the required items from. There can be multiple decision options, hereon referred to as degrees of freedom (DOF), because the same item types are often stored at multiple different locations. Hence, there may be no single right answer with various alternative choices being satisfactory; this is termed “satisficing” decisions, where satisfying a chosen aspiration level, rather than an optimal solution, is the focus (Gelman, 2010) as opposed to cases where there is only one optimal decision choice (see (Price and Shanks, 2011)). This type of decision is therefore referred to generally as a *multi-criteria, satisficing* decision. It is also analytical rather than judgemental, bargaining, or inspirational (Nutt, 1998; Shankaranarayanan et al., 2006).

Once a location is chosen, a picker picks the physical items and always returns to update the WMS, so the data is always current; warehouses are strict about data updates to ensure that data is always current. However, if a picker makes a mistake in the picking or placement of items, an inaccuracy will result when the WMS is updated because the pickers are unaware of their own mistake. There are various causes of inaccuracies in warehouses (Sarac et al., 2010), such as picking from the wrong location (DeHoratius and Raman, 2008) or misplacements (Lee and Özer, 2007; Rekik et al., 2008). This problem is compounded by the fact that there is no real indication of when a person will make these mistakes as they can occur anytime and at any location. Furthermore, it is not easy to detect the inaccuracies from the data alone, because when the locations to pick items from are selected from the WMS, the inaccurate data appears to be completely valid i.e. there is no obvious anomaly, such as more items at a location than is physically possible.

In this case, data tags that inform the WMS about which locations are inaccurate could be used to: 1. improve decision-making by enabling the WMS to choose the most accurate locations and hence help pickers avoid disruptions (such as being sent to locations that are empty) and, 2. help to find the inaccurate locations in the warehouse. However, to populate a data tag containing a true accuracy measurement is clearly problematic as there is no other source of this data, and hence continual inventory checks would be required to observe reality. Currently, warehouse organisations do use inventory checking as a way to ensure that the data in the WMS is accurate. However, it is difficult and inflexible for a warehouse to automate this process, and item tracking solutions such as Radio Frequency Identification (RFID) are expensive, inflexible, and suffer from inaccuracies themselves (Chan et al., 2012; Tu et al., 2018).

## 1.2 Characteristics of the general problem context

The above warehousing example is a particular decision-making case that exhibits the following general characteristics: 1. it is an analytical, multi-criteria, satisficing decision that can be improved by choosing better decision choices based on an understanding of the level of accuracy of the supporting data, 2. to measure accuracy is time-consuming, labour intensive, error prone and difficult/costly to automate, 3. a surrogate data source does not exist to support the measurement of accuracy, 4. data that is inaccurate

appears to be valid and does not exhibit any obvious anomalies, 5. the process to update the data does not correct the inaccuracy, 6. the data is kept current by regular updating of the data every time the real value changes, 7. the events that occur that cause the inaccuracies are known about in general, but it is not known when a particular instance of the event will cause an inaccuracy, 8. it is difficult or impossible to estimate the probabilities related to when data will become inaccurate.

Currently, there are no existing approaches that can provide an accuracy data tag to assist a decision-making problem with the above characteristics; Potential Problem Data Tagging aims to address this research gap. Section 2 describes the background research on accuracy, data tags and measuring accuracy automatically. Section 3 presents the Potential Problem Data Tagging approach, and Section 4 describes an application of the approach to the previously described warehousing example. Sections 5 and 6 describe and present the results of an evaluation of the approach using a simulation. Section 7 discusses the results and describes the limitations. Finally, section 8 concludes by discussing the implications of this work.

## 2 Background

This section defines accuracy, provides a review of the use of data tags, and presents the existing approaches related to measuring accuracy with reference to the above characteristics.

### 2.1 Defining accuracy

Accuracy has featured prominently as a key dimension in much of the data quality research (English, 2009; Fisher et al., 2009; Lee et al., 2002; Wand and Wang, 1996). The definition by Ballou and Pazer (1995) is an early example that covers a common theme of the misalignment between the data and reality, and is adopted in this research: *accuracy is “the degree to which the reported value is in conformance with the actual or true value.”* The term conformance alludes to the magnitude of the error (see the arguments presented in (Haegemans et al., 2016)) implying that it need not be a Boolean comparison between reality and data. The actual/true value implies that any assessment of accuracy requires an observation of this actual/true value in reality. Accuracy can be measured at the database level (Fisher et al., 2009), the

attribute level (English, 2003), and the record level (Suciu et al., 2011). For a good overview of accuracy definitions see (Fisher et al., 2009). We focus on measurement at the record level i.e. a record is accurate if all of its attribute values are accurate (Parssian et al., 2004).

## 2.2 Existing work on data tagging

For relational databases, tags can be used to convey the quality of data records (Suciu et al., 2011), columns (Price and Shanks, 2011), or individual values for various data quality dimensions (Wang et al., 1995). The values in the tags can vary between discrete-ordinal indicators of quality, such as representations analogous to traffic-light signals (Price and Shanks, 2011; Watts et al., 2009) to continuous numerical indicators (Suciu et al., 2011).

Various multi-attribute decision-making tasks have been considered in the data tag research, including residential apartment selection (Chengalur-Smith et al., 1999; Fisher et al., 2003; Price and Shanks, 2011), ranking of firms according to their financial health (Moges et al., 2016), ranking of job alternatives (Fisher et al., 2003), and allocation of budget among multiple types of advertising (Watts et al., 2009). These works all assume, and rely on, a data tag for accuracy being available.

## 2.3 Existing approaches to measure accuracy

The following sections describe the different techniques and metrics that can measure accuracy. Approaches that address the quality dimension of currency are also reviewed, due to the similarity of currency to accuracy (see the discussion in (Heinrich and Klier, 2015)).

### 2.3.1 Valid but inaccurate data

Validation rules can be used to check whether data is likely to be inaccurate, such as that a social security number must consist of nine numeric digits (Olson, 2003), and it is considered to be inaccurate otherwise. Inaccuracies can be inferred using dependencies (Sessions and Valtorta, 2009) or rules relating to multiple data columns (Woodall et al., 2014). Validation rules can also use temporal aspects to infer inaccuracies (Maydanchik, 2007). Anomaly/outlier detection solutions can be used to find patterns in data that do not

conform to expected normal behaviour (Chandola et al., 2009; Hipp et al., 2007), which can also be considered a type of validity check. However, despite the references to accuracy, English suggests that many of these types of problems are validity checking (English, 2003); hence the data could be both valid but inaccurate. For example, a person's social security number recorded in a database could contain nine numeric digits and still not be the social security number of that person. Hence, these types of approaches are not suitable to address the problem context in this paper because of the characteristic that data appears to be valid but can still be inaccurate.

### 2.3.2 Unavailable or inaccurate surrogate sources

A surrogate source of data may be available from which the data can be checked against (English, 2009). Online sources can be used, although the challenge is the extent to which they accurately represent reality itself (Lukyanenko et al., 2014). Also, data redundancy between systems can be used to cross-check data accuracy (Han et al., 2010). However, surrogate sources cannot be used to detect inaccuracies when a surrogate source does not exist and/or is not correct itself (Wechsler and Even, 2012), which is the case for the characteristics of the problem context in this research.

### 2.3.3 Statistical analysis and currency

Inaccuracies can be estimated using statistical analysis of a series of measurements to determine estimation errors (Joglekar et al., 2013), or performing a sample of observations and then estimating the overall level of accuracy of a set of records (Motro and Rakov, 1997; Sessions and Valtorta, 2009). Other statistical work observes currency of data that can become inaccurate if the real-world counterpart changes over time and the stored data values are not updated (Heinrich and Klier, 2015; Wechsler and Even, 2012). Markov models have been used to model the probabilities that attribute values change, and can estimate when data will become inaccurate (Wechsler and Even, 2012; Zak and Even, 2017). Various useful currency metrics have been developed using probability theory (Heinrich et al., 2009; Heinrich and Hristova, 2016; Heinrich and Klier, 2011, 2015).



The approach presented in this paper builds on the foundations of the currency approaches and makes three important changes: 1. removing the need to provide estimates of the transition probabilities for data values changes, 2. shifting the focus from the passage of time to the occurrence of events, and 3. the ability to operate with or without probability estimates.

The existing currency models rely on being able to estimate the transition probabilities for data value changes; hence, they are ideally suited to examples where these can be estimated, such as marital status (Heinrich and Hristova, 2016), or the changing status of a process (Zak and Even, 2017). However, contrast this to the quantity of items at a warehouse location transitioning from 4 to 5 items, and what the associated probability is. There are two challenges with this: 1. it is difficult to determine what the transition probabilities should be and 2. there are a very large number of transitions that require probability estimates. Hence, the key adaptations for the context in this paper is that the equivalent of Markov model states now refer to whether the data is accurate or inaccurate rather than having states for each individual data value. Then the transition probability becomes an estimate of whether the data will become inaccurate or not. And this can be simplified again to: a tally of the number of times data has been exposed to an event that can cause it to become inaccurate. Hence, in cases where it is difficult to estimate the probabilities, the information that must be known is only when events occur that may cause data to become inaccurate. Note that the shift from time to events is important in the case when it is known that both a particular event may cause an inaccuracy and when this event will occur (as in the warehouse example: the picking event may cause an inaccuracy and it is known when it occurs). Data that over a long period of time has been exposed to no events that may cause it to be inaccurate is less likely to be inaccurate than data that has been exposed to many of the events (regardless of the time period).

### 3 Potential Problem Data Tagging

A Potential Problem Data Tag is a recording of the number of times data has been exposed to an event that can cause it to become inaccurate. The event may not always cause the inaccuracy; hence, it does not

produce an actual measure of accuracy, and must not be treated as such. However, the tag can be used to determine which data records are likely to be more accurate or less accurate than one another; this can be considered to be “relative accuracy” (Fan and Geerts, 2012, p. 5) between records in a database (note that this paper focusses on database records, further work would be required to extend the approach beyond this). These types of tag can be used to assist with operational decision making as shown in Figure 1, which shows a database and front-end software application that is used to make decisions that are actioned in the physical operations.

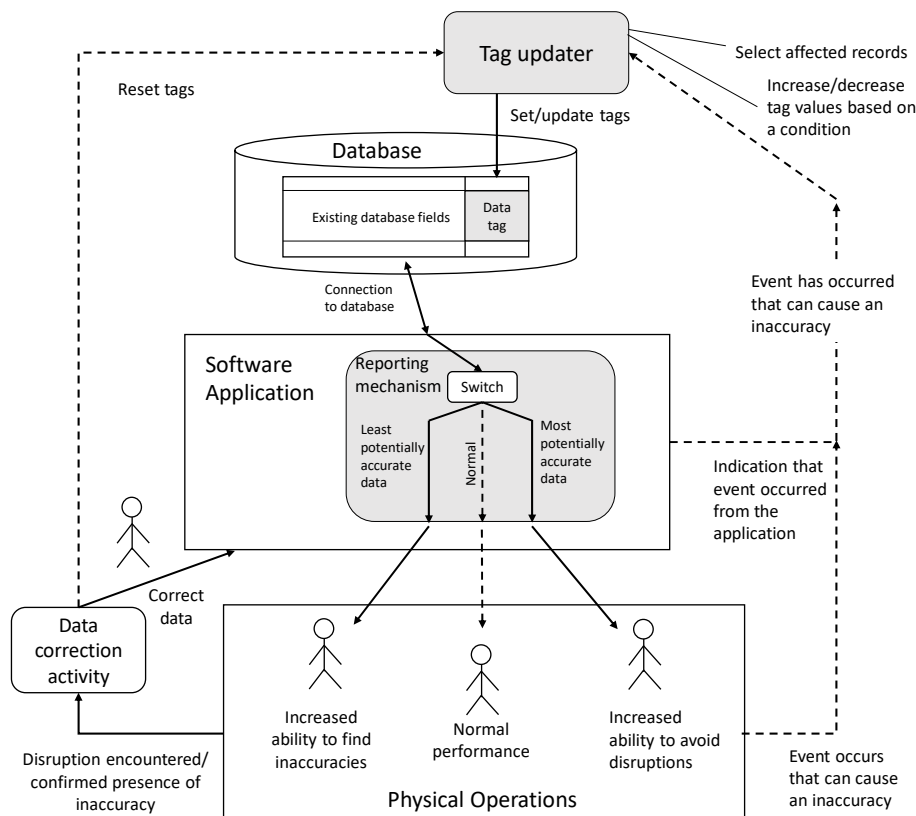


Figure 1: Illustration of the Potential Problem Data Tagging Approach

The tags can be used to select either the least or most potentially accurate data from the database, which can then be used within the operations to find inaccuracies or avoid disruptions respectively. The increased ability to find inaccuracies can be used to support data correction activities, whilst avoiding disruptions can lead to higher operational performance. This capability is dynamic and can be switched at will, which can therefore align with the demands of the operations: such as in cases where time is critical and avoiding

disruptions is required, to the opposite case where there are slack periods that allow staff to spend more time improving the quality of the data by finding inaccuracies. The switch is achieved by filtering the database records depending on the tag values, and the reporting mechanism of the application is the point at which this can be implemented (see Figure 1).

A tag can be incorporated as an additional field in the database, and the tag updater component is responsible for setting/updating the values of the tags for the relevant data. The tag values must consist of any set of symbols that can produce an ordinal scale. For example, a convenient choice is a numeric value with zero indicating an accurate record, and values between zero and one indicating increasingly likely to be an inaccurate record, eventually with a value of 1 indicating a definitely inaccurate record. The tag values are incremented based on events occurring that can cause inaccuracies to be inserted into the data. An event is something that happens in the physical operations of an organisation (e.g. a picker going to pick items in a warehouse, or the maintenance of a machine, etc.) that can cause an inaccuracy. Ideally an event can be observed automatically; however, if this is not possible, there may be an indication that the event occurred from the software application (an engineer entering a maintenance record, for example, indicates that maintenance of a machine occurred). Once the event occurs, it may only produce an inaccuracy in certain records in the database. In many cases the selection of which records to tag will be clear by the context, and it is especially clear when the event is informed by the software application. For instance, when entering a new maintenance record for a machine, the inaccuracy is between that record and the machine (rather than for any other machine or any other database record). The tag updater component selects the affected records and updates their tag values when the event occurs. The tag updater must also reset the tag values after any data correction activity to indicate that the relevant data is now accurate. Finally, under certain conditions an event may be more or less likely to introduce an inaccuracy, and therefore the tag updater component should be able to modify the tags to reflect this increase/decrease in likelihood.

### 3.1 Setting the tag values.

There are two distinct cases that determine how to set tag values: if the probability of an inaccuracy is not known and cannot be easily estimated, versus if it is known/can be estimated. In the former case, it is sufficient to use a tally of event occurrences for each data record, which is updated each time the event occurs, this produces the required ordinal scale in a simple way. In the latter case, Table 1 shows one example of how to use a probability estimate for setting and updating tag values that also produces an ordinal scale. Note that this is a convenient strategy because the updates to the tag will tend to 1, but will never reach it, meaning that the value of 1 can be reserved for definitely inaccurate values. This construction satisfies the minimum/maximum requirements for metrics (Heinrich et al., 2018). In the table, *currentValue* is the current value of the tag and EP is the “Expected Probability” of the event that causes an inaccuracy in the data record. EP can be estimated using observations of prior events and by calculating: *the number of inaccuracies observed/the total number of events observed*. The formula in Table 1 is essentially the probability of the union of two events (for non-mutually exclusive events); it is equivalent to the probability that the inaccuracy occurred for the current event or the previous events or both. In this case, we assume that EP is constant for all events, the events are assumed to be independent, and also possesses the memoryless property because no information is gained from previous events. Note that these tag values align with the formalisms used in probabilistic databases (Suciu et al., 2011) and decision making (Heinrich and Klier, 2015). Furthermore, the value of EP can be set dynamically based on conditions related to the event; for example, it could be increased under conditions where the event is more likely to lead to an inaccuracy, and vice versa (an example of this is shown in section 4 for two different EP values which are set dynamically).

<b>When to set/update the tag</b>	Starting value (or whenever the data is confirmed as being correct)	The first time the event occurs	All subsequent times when the event occurs	Whenever the data is confirmed to be definitely inaccurate
<b>Value to set in the tag</b>	0	EP	$(\text{currentValue} + \text{EP}) - (\text{currentValue} * \text{EP})$	1

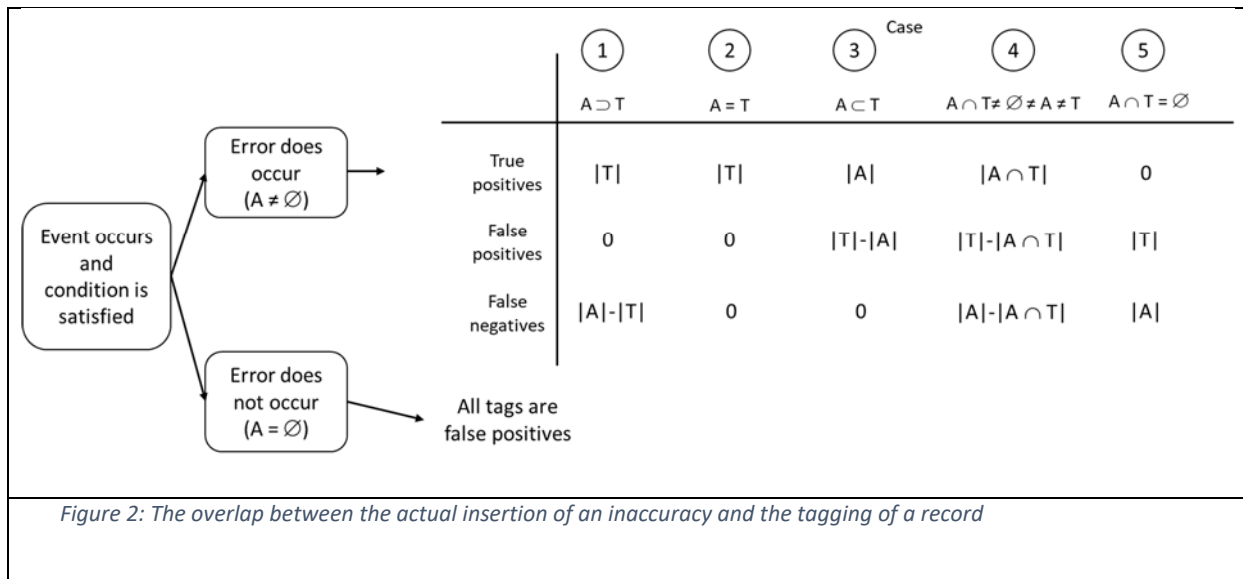
Table 1: When and how to update the tags (EP is the expected probability)

### 3.2 Assumptions for use

The following assumptions must be satisfied for Potential Problem Data Tagging to be applicable to a problem and to potentially yield a benefit: 1. **DOF > 1**: the decision/task must have more than one DOF 2. **tag correctness**: when the event occurs, the tag values correctly tag an inaccuracy in at least one case (i.e. there is at least one true positive tag), 3. **record differentiation**: when the event occurs, at least 1 record is tagged and not all records associated with the DOF are tagged.

Assumption 1 must hold true otherwise it is not possible to choose alternative courses of action for the decision, and it would not be a satisficing decision (cf. section 1.2). Assumption 2 concerns the ability to correctly tag a record when it contains an inaccuracy. The challenge is that a) the choice of which record(s) to tag needs to be correct (i.e. it is possible to tag the wrong record(s)), and b) an accuracy may not always be inserted into a record because the event may not always result in an error that leads to an inaccuracy. These situations lead to tags being false negatives and false positives respectively. All of the possible situations are summarised in Figure 2. In Figure 2, 'A' is the set of records actually containing an inaccuracy caused by the event, and 'T' is the set of records that have been tagged as containing an inaccuracy for the event. Note that  $|A|$  refers to the number of actual inaccuracies that have been inserted into the data for a given event. For cases 1, 2, 3 and 4, there is the possibility that the approach will provide a benefit; case 2 being the best case where all the tags correctly tag all of the records containing the actual inaccuracies. The approach does not provide any benefit in case 5 because there are no tags that are true positives.

Assumption 3 (i.e. record differentiation) must be satisfied because otherwise there is no differentiation between records that helps to indicate which records are potentially more likely to be accurate than others. The same is true if DOF records are tagged. Note that these three assumptions need not always be satisfied (for every event occurrence) because the DOF, true positives etc. can all change for each different instance of an event.



## 4 Application to decision making in a warehouse

This section illustrates how to apply Potential Problem Data Tagging to the warehousing decision-making example presented in section 1.1. where the data tags are used to improve decision-making by enabling the WMS to choose the most inaccurate/accurate locations and hence help pickers either find inaccuracies or avoid disruptions. Inaccuracies in the data can lead to the following disruptions: D1) pickers are sent to warehouse locations that contain too few items (or are empty), D2) pickers are sent to locations that contain the wrong item types, D3) the WMS rejects orders that the warehouse can fulfil assuming there are not enough items in the warehouse, D4) the WMS accepts orders that the warehouse cannot fulfil assuming there are enough items in the warehouse. D1 and D2 are caused by the aforementioned misplacements or picking from the wrong location. Misplacements occur when items are put back into the wrong location either on top of similar items or into empty locations. Picking from the wrong location is when the picker takes the items from the wrong location. In later picks, if it is observed that there is nothing in a particular location when items are expected to be there, then the WMS is updated to reflect reality i.e. the location is set to zero quantity (so that pickers are not sent again to that location to face the disruption again). However, this has two side effects: firstly, the actual quantity in the warehouse is reduced so it is possible to reject orders which can actually be fulfilled (D3); secondly, the inaccuracy is not entirely solved because

there is another location where the items actually are that will remain inaccurate. The opposite situation can occur when items are found in the warehouse that are not in the WMS, and hence, disruption D4 can occur. Not all inaccuracies necessarily lead to operational disruptions, however. For example, if the WMS records that there are too many items in a location, then the picker will not face a disruption because there will be enough for them to pick. Also, it is unlikely they will even notice the data inaccuracy as a) they are not aware of the value stored in the WMS while performing the picking task, b) it is impractical to always count the items lying in a storage location without spending considerable amounts of time.

The following sections describe how Potential Problem Data Tagging was implemented for the experimental evaluation, to help select the most accurate or inaccurate locations for the above warehousing example.

#### 4.1 Event selection

The printing of a picking list by the WMS was selected as a convenient event that acts as a surrogate for the insertion of an inaccuracy. In reality, when a picking list is printed, it is handed to a picker whom, while physically picking the items, could misreplace or pick the wrong items. Due to the difficulty in determining exactly which events will lead to a misreplacement or incorrect picking error, the solution always records that the error could have been made.

#### 4.2 Selection of records to tag

There are two instances of inaccuracies that result from misplacements and picking from the wrong location: 1. in the record about the *expected* picking location, and 2. in the record about the *actual* picking location (where items have been actually taken from or misplaced to). The “expected” record is known, and the “actual” record is unknown. For this reason, only the expected record was tagged. Also, if a misreplacement error does occur, then the expected record is guaranteed to be inaccurate. This solution therefore fits case 1 in Figure 2 as it will generate one true positive and one false negative. Hence, the second assumption (tag correctness), which indicates whether the solution can provide a benefit, is satisfied.

The final assumption (record differentiation) is also satisfied as the solution will always tag 1 record for an event. The first assumption is satisfied as long as the warehouse contains the same type of items in multiple locations, which therefore provides  $DOF > 1$ .

### 4.3 The tag and tag updater component

The tag was implemented as an extra field in the WMS database. To demonstrate the case where EP is not known (e.g. in some warehouses there is no data on the errors), a simple tally was used starting at 0 and incremented by 1 each time the event occurred. To demonstrate the case when EP is known, the tag value of EP was set to 0.02 (based on observations of inaccuracies from historical inventory checks in a real warehouse where data was available). Furthermore, an additional case was also evaluated where according to observations in reality, a misplacement is more likely to occur when there are empty locations nearby to the picking location (within approximately +/- 3 locations). Hence, when this condition is true, the value of EP was increased to 0.05; a conditional check for nearby empty locations in the WMS was carried out by the tag updater component, and if true, the increased EP value was used, otherwise the default value of 0.02 was used. The tag updater component was implemented in Java, and it updates the tag values (as shown in Table 1) using SQL queries issued to the WMS database. It updates the tags when: 1. the picking list is printed—the tag is updated to indicate that an inaccuracy may have been inserted into the data, 2. after a data correction activity (inventory check) it is set to zero, 3. after a picker encounters a disruption thus confirming that there is an inaccuracy—the tag is set to 0 because the WMS is updated with the correct data for the location, and 4. after a replenishment, it is set to zero for the replenished location.

### 4.4 The reporting mechanism switch

The WMS was configured to return the locations with the required item type and sufficient quantity and also with either the lowest tag value (to avoid disruptions) or with the highest tag value (to find inaccuracies). Importantly, the values within the tags are, therefore, never shown to the pickers. Table 2 shows an example set of tag values that arise after using this mechanism (and in comparison to the simple



tally value in the final column) to update the tags in the simulated WMS database. Note that locations 2, 3 and 4 are within 3 locations of an empty location, and so they apply an EP value of 0.05 rather than 0.02. The table shows the state of the WMS after 6 picks starting with an inventory check (i.e. all data was accurate). Items have been picked from location 1 once, location 3 twice, and location 4 three times; the tag values reflect this. Hence, location 4 is most likely to contain an inaccurate value. Location 2 is the best location to pick item type 'A' from in order to avoid disruptions. In the warehousing example, a picker could aim to "avoid" disruptions by being prompted to pick from storage locations that are most likely to contain accurate data. For example, if 10 items of type 'A' product can be picked by either storage location 1 or 2 (according to the WMS), then the "avoid" configuration would choose the location that is more likely to include accurate data in the WMS (i.e. location 2). The picker could aim to find inaccuracies by being prompted to pick from storage locations that are most likely to contain inaccurate data. In this way, visiting a storage location to perform a picking task is more likely to lead to a disruption that will allow the picker to notice the data inaccuracy too (e.g. a picker arrives at a location where there are too few items and so they know that there is an inaccuracy because they cannot complete their task). Whether it is the wrong quantity, item type or both, this can be reported back to the warehouse manager to update the WMS with the data reflecting the actual observation, and the tag is set to 0. At this point the manager knows that there is definitely another inaccuracy somewhere in the warehouse where the remaining missing items are. The warehouse manager may then decide to start an investigation to find the missing items (during picking activities) or could document the error and determine if an inventory check is needed (warehouses typically use the number of encountered inaccuracies as one way of triggering an inventory check).

Location	Item type	Item quantity	Tag (Using EP)	Tag (Tally)
1	A	30	0.02	1
2	A	20	0	0
3	B	15	0.0975	2
4	B	4	0.14265	3
5	-	0	0	0

*Table 2: an example of the tag values for a database containing information on locations of items in a warehouse*

In order to better understand the use of the two configurations, it is important here to differentiate again between data inaccuracies and operational disruptions. As noted previously, inaccuracies do not always lead to disruptions unless the picker is asked to perform a task that they cannot complete due to the inaccuracy itself, and they will only know to report back when they face a disruption; and the tag is set back to 0 for the relevant location because the WMS is corrected. However, this does not correct the inaccuracy entirely because there is still a second instance of the inaccuracy in the warehouse where the missing items are located.

## 5 Experimental Evaluation

This section describes the experimental simulation that Potential Problem Data Tagging was evaluated on, based on the previously described warehousing picking scenario. The following hypotheses were tested in the avoid and find configurations respectively:

- AH0: There is no difference between the number of disruptions encountered when using Potential Problem Data Tagging to avoid disruptions compared to using a WMS as normal.
- AH1: Fewer disruptions are encountered when using Potential Problem Data Tagging to avoid disruptions compared to using a WMS as normal
- FH0: There is no difference between the number of disruptions encountered when using Potential Problem Data Tagging to find disruptions compared to using a WMS as normal
- FH1: More disruptions are encountered when using Potential Problem Data Tagging to find disruptions compared to using a WMS as normal

These hypotheses were analysed under different cases including: using simple tally values for the tags versus using EP set to 0.02, replenishing items versus not replenishing items, and use of two different EP values of 0.02 as the default and 0.05 when a misplacement is more likely to occur. Potential Problem Data Tagging avoid and find configurations do not eliminate disruptions, but shift them until later or earlier

picking tasks respectively; the hypotheses were therefore also compared for 1 to 100 picks against 1 to 200 picks.

### 5.1 Simulating the warehousing problem

To test the aforementioned hypotheses, a simulation methodology was chosen. Typically, simulations have advantages over other research methods if a study concentrates on mechanisms and processes (Gilbert and Troitzsch, 2005) like the warehouse picking operation studied here. Indeed, simulation approaches have been followed often in process analyses and are a commonly used tool to investigate warehouse operations (Faria and Reis, 2015; Giannikas et al., 2017; Pedrielli et al., 2016).

The warehousing decision-making problem was implemented in software (using Java and MySQL) with one database representing the WMS and another, the physical warehouse. Operations on the warehouse and WMS were modelled by querying and updating the data in both databases, and inaccuracies could be checked by comparing the two. In the simulation, orders arrive at the warehouse and the WMS chooses the locations to pick items from that satisfy the orders. While picking the items, pickers can make the following mistakes, which result in inaccuracies, and can subsequently lead to disruptions: 1. misreplacement (occurring when, after taking all the items off a shelf, pickers put the remaining quantity back into the wrong empty location), 2. picking from the wrong location (occurring when a picker takes the items from a different location to the one that is specified on the picking list). As a result of these two mistakes, the following three disruptions can be encountered in the simulation: D1 (empty locations), D2 (wrong items) and D3 (WMS rejects orders that the warehouse can fulfil). Disruption D4 (WMS accepts orders that the warehouse cannot fulfil) is not faced in the simulation because the orders were chosen so that the warehouse can always satisfy the order. In the normal case, referred to in the hypotheses, there is no indication of priority of which locations are inaccurate or not (because as noted previously in section 2, there is no current solution that can provide this). The normal case, however, does correct some inaccuracies in a sensible way, including when the picker does not find enough items to pick or finds the wrong item type, they return and update the WMS to reflect what they observed. This sensible correction of inaccuracies is also done in the same way for

the avoid and find configurations in order that the comparison is fair. Note that this type of correction does not solve the inaccuracy entirely because there is another inaccuracy instance (described at the end of section 4) where the items have been moved to, which requires a more thorough check to find and correct. Due to the complexity of this check, it was not implemented.

## 5.2 Experimental configuration and measurement

The experiments were run over 500 trials of 200 picking tasks for each of the three configurations: normal, avoid and find. The simulation explored three different error insertion rates (1%, 5%, and 20%) and three different DOF (2, 12, 60). The hypotheses were tested for all of these combinations. A 20% error rate is a very high value for warehouses, 1% is a low value, and 5% represents a typical value. A DOF of 2 is the minimum possible, and 60 is the average DOF that results from having all items the same in the warehouse simulation; 12 DOF represents a typical normal value that results from having 5 different items in the warehouse.

Variable	2 DOF	12 DOF	60 DOF
Warehouse locations	100 (each is unique and only one item type is allowed in one location)		
Initial configuration of the warehouse and WMS	No inaccuracies, random insertion of item type, random insertion of quantity, random placement into locations. 5 different initial configurations were created and used as blocking factors in the trials.		
Empty locations	20 (randomly distributed)		
Number of different item types in the warehouse	Random between 1 and 50	Random between 1 and 5	1 (all items are the same type)
Quantity for each item type in the warehouse	Random between 50 and 100	Random between 1 and 100	Random between 1 and 100
Orders (picking tasks) per trial	200 per trial. The order contains one random product type and quantity. Orders were chosen deliberately so that the warehouse can always satisfy the order. (quantity and item type) by selecting an existing item type and not exceeding the quantity in the warehouse.		
Order quantity	Random between 1 and 5 (or maximum in the warehouse)	Random between 1 and 20 (or maximum in the warehouse)	Random between 1 and 20 (or maximum in the warehouse)
Insertion of errors (misalignments)	1% error rate (2 errors), 5% error rate (10 errors), and 20% error rate (40 errors) Attempted at random picking tasks. Some may not be successful if there is nothing to pick for that task or there are no locations available to misreplace items. The		

	first location (with the same item type or empty) given by the WMS is chosen to misreplace the items or pick from the wrong location.
Types of errors inserted	Misreplacement and pick from wrong location. Always split 50% each per trial. For misreplacements, they are 4x more likely to occur if there are empty locations are within 3 locations of the pick location.
Satisfying orders	Pickers only pick from single locations that have enough quantity available to satisfy the order.
Picking policy	The actual location from which items are picked is chosen as the first location in the list generated by the WMS.
Inaccuracy correction	If an inaccuracy is found, it is corrected only by updating the entry in the WMS. The other location where the items have been misplaced to was not updated in the WMS (as in reality, this would need to be physically found).

*Table 3: Parameterisation of the experimental variables*

The experimental variables and their parameterisation are summarised in Table 3, with the differences shown for each DOF. All other factors were randomised (see Table 3). The disruptions were measured by incrementing a counter when the picker faced a disruption. A picker may pick from a location with an inaccuracy and not know about it if there are more than enough items to pick; to reflect reality, these instances were not counted as disruptions in the simulation.

## 6 Simulation Results

This section presents the results of the simulation for avoiding disruptions and finding inaccuracies versus a normal operation of a WMS. A student's T test (Montgomery, 2012) was used to identify any significant differences between the mean number of disruptions encountered.

### 6.1 Using a tally or a static setting of EP values

The first results shown are for the tally values and EP set to 0.02, which were confirmed as being the same after simulation executions of each. This is because having EP set to a single value and using the formula in Table 1 is logically the same as using a tally; the only difference being in the symbols used with the EP values providing a more useful interpretable probability value as described in section 3.1.

### 6.1.1 Avoidance of disruptions

Figure 3 illustrates the mean values for all cases tested. In all cases (including for 200 and 100 picks) Potential Problem Data Tagging is significantly different from the normal at the 99.9% level leading to reject the null hypothesis (AHO) in these cases; the only exceptions being for 200 picks at the 1% error rate with 2 and 12 DOF, which are not significant, and the 5% error rate with 12 DOF, which is only significant at the 95% confidence level.

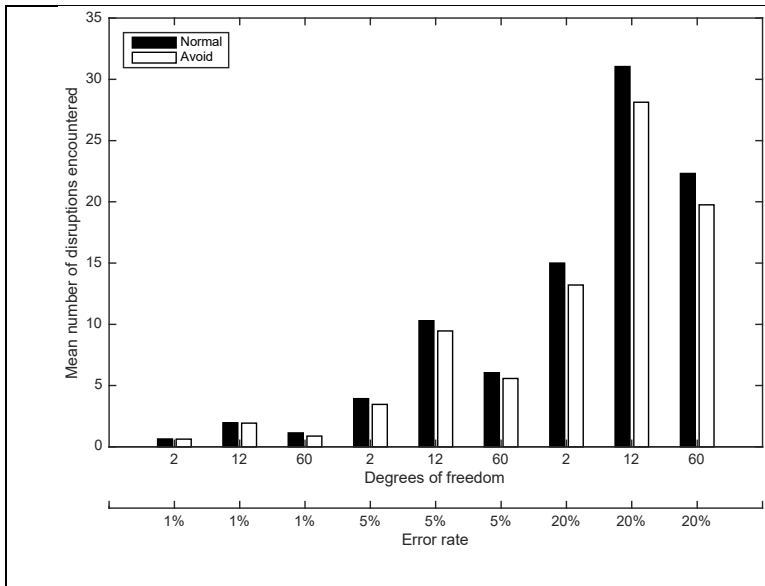


Figure 3a: Configuration = Avoid, Picks = 200

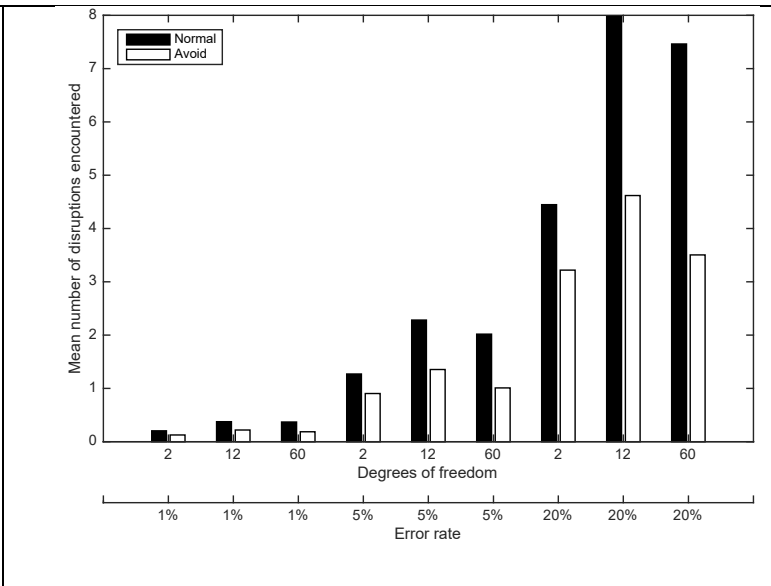


Figure 3b: Configuration = Avoid, Picks = 100

Figure 3: Mean Number of Disruptions Encountered for Potential Problem Data Tagging (avoid case) versus the Normal Operation of a WMS (Tally and EP = 0.02 case).

### 6.1.2 Finding inaccuracies

When attempting to find inaccuracies, all results are significant at the 99.9% level, and the 1% error rate with 12 DOF is significant at the 95% level. Figure 4 shows the comparison of the approaches against the normal operation of a WMS. The results show that in all cases save one, Potential Problem Data Tagging is significantly different from the normal operation of a WMS at the 99.9% confidence level. The other case (1% error rate with 12 DOF) is still significant, but at the 95% confidence level instead of 99.9%. For the first 100 picks, the 1% error rate with 12 DOF also becomes significantly different at 99.9% confidence. This

indicates that the find performance also degrades over time. This can also be seen by observing the larger differences between the mean number of inaccuracies found between the cases in Figure 4b compared to the smaller differences in Figure 4a. The null hypotheses (FH0) are rejected for all cases when attempting to find inaccuracies using Potential Problem Data Tagging.

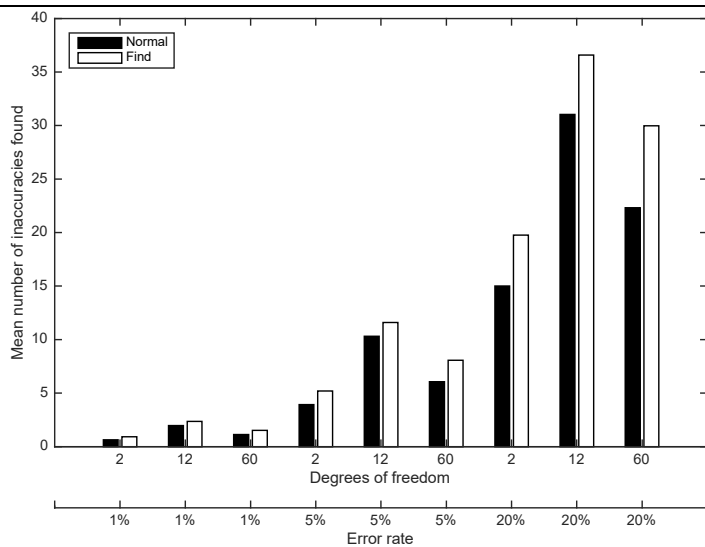


Figure 4a: Configuration = Find, Picks = 200

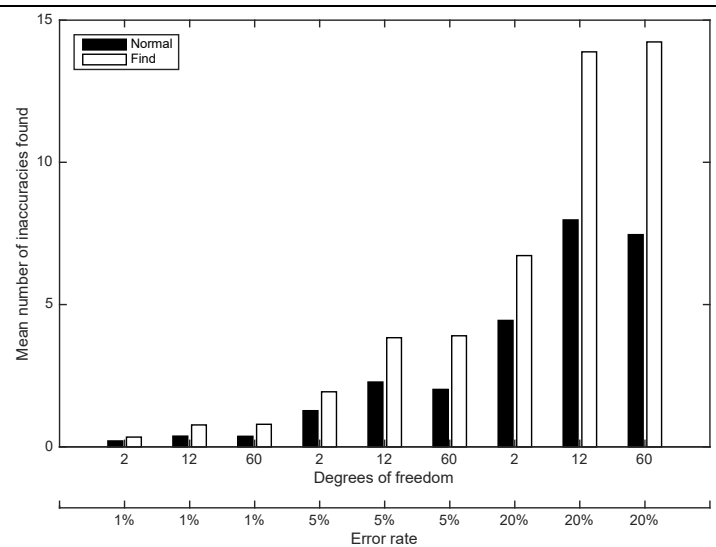


Figure 4b: Configuration = Find, Picks = 100

Figure 4: Mean Number of Inaccuracies Found for Potential Problem Data Tagging (find case) versus the Normal Operation of a WMS (Tally and EP = 0.02 case).

A confidence interval estimates analysis indicates that it is possible to find from approximately 7 to almost 8 more inaccuracies over 200 picks compared to the normal situation for the 20% error rate with 60 DOF. This equates to finding between 17.5% and 20% of the 40 inaccuracies inserted into the data in 200 picks. This falls to approximately 4 to 6 inaccuracies for the 20% error rate with 2 and 12 DOF. When considering only the first 100 picks for the 20% error rate with 2 DOF, the number of inaccuracies found falls to approximately 2. It also falls for the 20% error rate with 60 DOF to between 6 and 7 inaccuracies. However, for the 20% error rate with 12 DOF the confidence interval itself increases from (5.58, 6.23) for the first 100 picks to (4.54, 6.55) for all 200 picks.

## 6.2 Using dynamic setting of EP values

Secondly, the results are shown for dynamically setting the EP value based on the context (using a default value of 0.02 and increased to 0.05 when it is more likely that an inaccuracy will be inserted); these results are also presented both with and without replenishment of items in the warehouse.

### 6.2.1 Without replenishment:

When dynamically setting the EP value based on the context to avoid disruptions, the only cases that were not significant were for DOF of 2 and 12, for the 1% error rate for 200 picks. For the 100 picks case only the DOF of 2 for 1% was not significant. All other cases were very significant over the 99.9% level. For finding inaccuracies, all cases were very significant except the 1% error rate for 12 DOF for 200 picks, which was significant at the 99% level; and for 100 picks the 1% error rate for 2 DOF was significant at the same level.

The results for these cases are shown in Figure 5 and Figure 6.

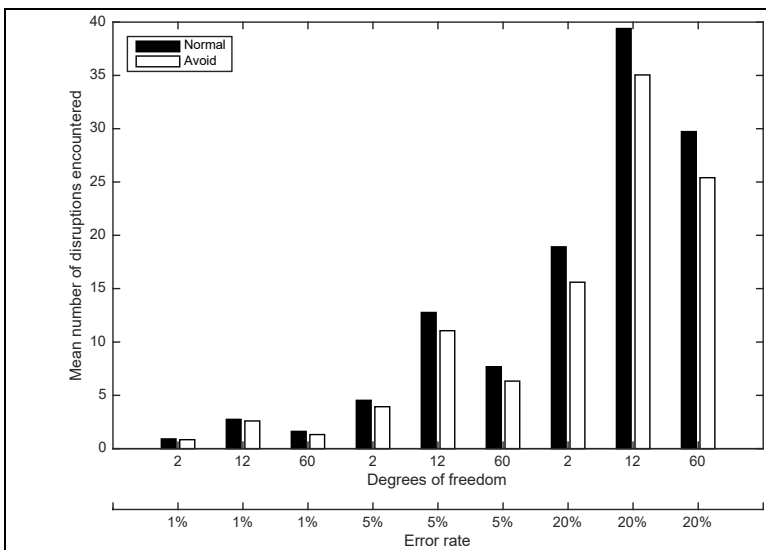


Figure 5a: Configuration = Avoid, Picks = 200

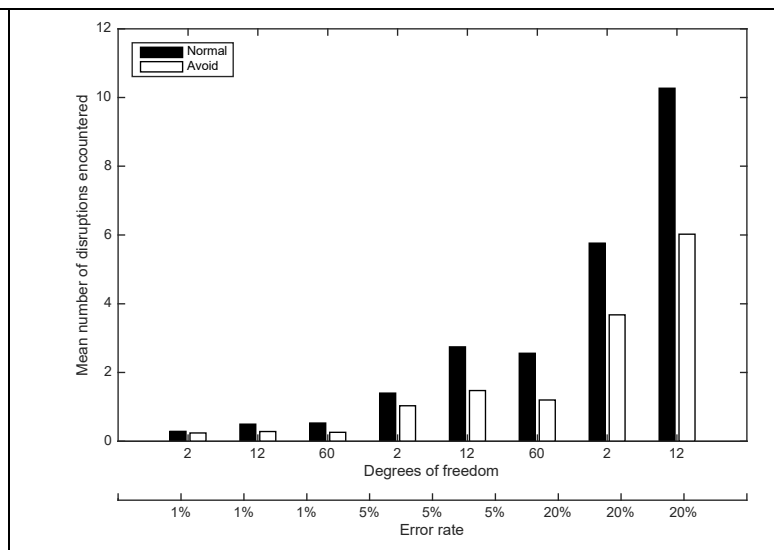


Figure 5b: Configuration = Avoid, Picks = 100

Figure 5: Mean Number of Disruptions Encountered for Potential Problem Data Tagging (avoid case) versus the Normal Operation

of a WMS (dynamic case without replenishment)



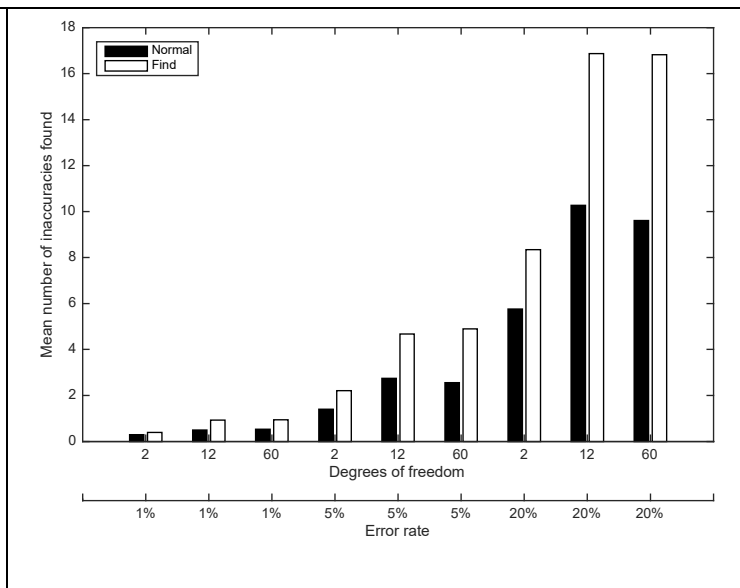
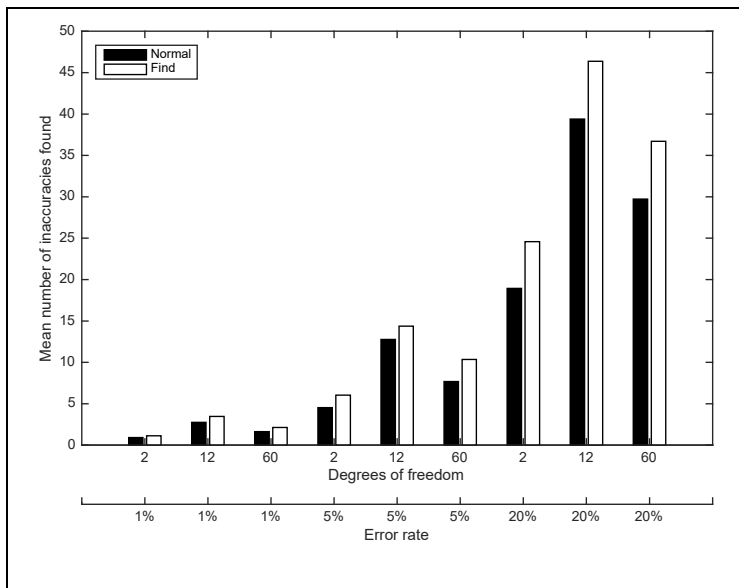


Figure 6a: Configuration = find, Picks = 200

Figure 6b: Configuration = find, Picks = 100

Figure 6: Mean Number of Inaccuracies Found for Potential Problem Data Tagging (find case) versus the Normal Operation of a WMS (dynamic case without replenishment)

### 6.2.2 With replenishment:

The previous results were collected without replenishment of items in the simulation. The following results included items being replenished back to a quantity of 100 at 75 and 150 picks for locations where the quantity was lower than 10 items. For avoiding disruptions, all cases are significant at the 99.9% level.

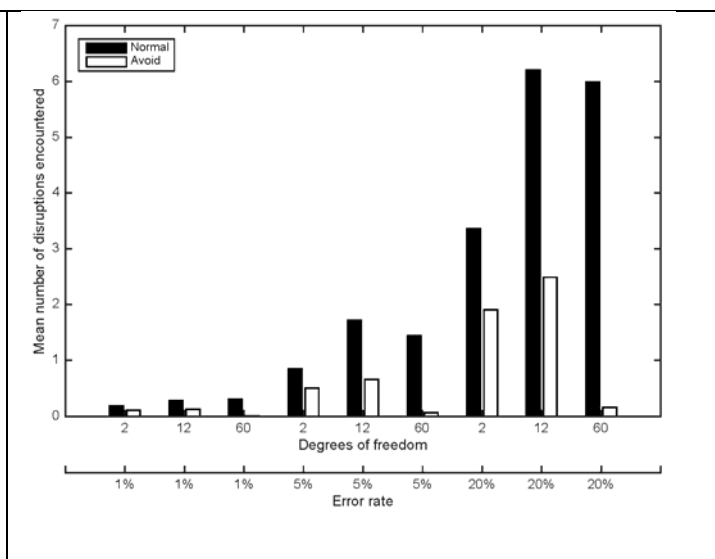
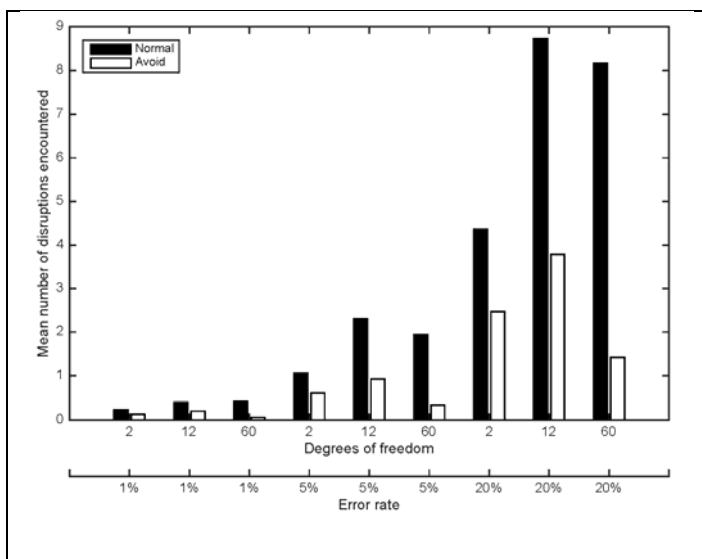


Figure 7a: Configuration = Avoid, Picks = 200

Figure 7b: Configuration = Avoid, Picks = 100

Figure 7: Mean Number of Disruptions Encountered for Potential Problem Data Tagging (avoid case) versus the Normal Operation of a WMS (dynamic case with replenishment)

When attempting to find inaccuracies, the results show that in all cases save one, Potential Problem Data Tagging is significantly different from the normal operation of a WMS at the 99.9% confidence level (see Figure 8a). This is also true for the first 100 picks (see Figure 8b). For both 100 and 200 picks, the other case (1% error rate with 2 DOF) is still significant, but only at the 95% confidence level. The null hypotheses (FH0) are therefore rejected for all cases.

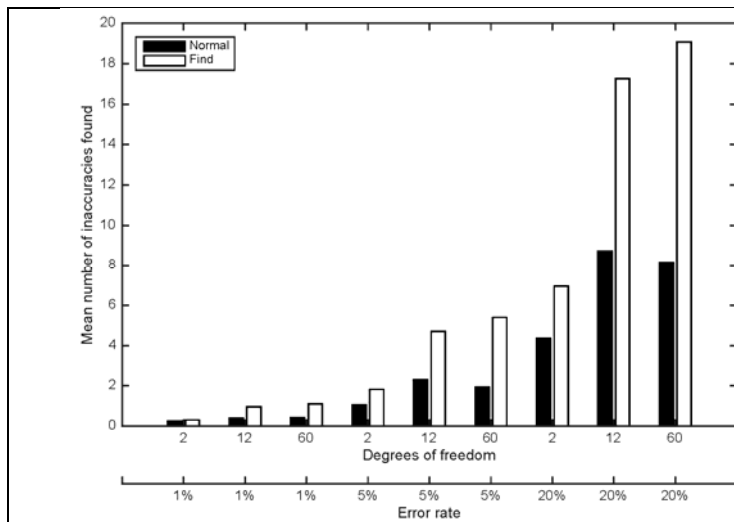


Figure 8a: Configuration = Find, Picks = 200

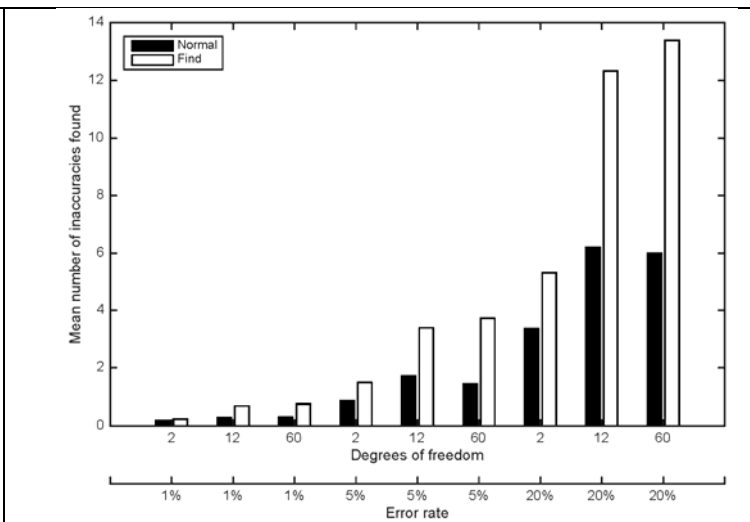


Figure 8b: Configuration = Find, Picks = 100

Figure 8: Mean Number of Inaccuracies Found for Potential Problem Data Tagging (find case) versus the Normal Operation of a WMS (dynamic case with replenishment)

## 7 Analysis and Discussion

For both the avoid and find configurations, the performance against the normal improves as the error rate increases. For the avoid 200 picks case, the confidence interval estimates increase monotonically as the error rate increases. This is also the case for the find configuration save for 20% error rate with 2 DOF case, which does not differ drastically and is likely due to the inherent variation in the simulation. As the error rate increases, the false positive tags (that result because the error does not occur) are replaced by true positives

(see “error does not occur” and case 1 in Figure 2). Hence, the tags become more accurate as the error rate increases. The improvement can also be explained partly due to the fact that, when avoiding disruptions, the normal case performs comparatively well when there are few errors. This is because there is a high chance a disruption will be avoided since there are many storage locations that an item can be picked from without causing a disruption.

In general, the find case shows greater performance than the avoid case, which can be seen by the larger number of inaccuracies that can be found (especially for higher error rates with higher DOF) compared to the smaller number of disruptions avoided in the avoid case. There are a number of reasons why find outperforms the avoid case: in the find configuration the same storage locations are picked from for different orders until their quantity runs out (the performance is reduced slightly due to the replenishments). As a result, the find configuration tends to group the errors into a smaller number of locations that have a much higher chance of containing an inaccuracy. This is not true in the normal case, where the inaccuracies are more uniformly spread between locations. This phenomenon also makes it more likely that a location with an inaccuracy from a previous picking task is accidentally tagged again as having an inaccuracy. Therefore, some of the false negatives become true positives, which improves the performance.

Regarding DOF, one would expect the performance to improve as it increases because there is more opportunity to choose different storage locations that are more or less likely to be inaccurate. This is confirmed in all cases where the confidence interval estimates show that as the DOF increase the performance of the find and avoid against the normal improves. The results show that this effect is larger as the error rate increases.

The confidence interval estimates analysis indicates that both the avoid and find configurations for 200 picks very slightly outperform 100 picks. This outperformance is more evident for the 20% error rate compared to 1% and 5%. In the find case this again improves. This is likely due to the solution always outperforming the normal and having more time to avoid disruptions or find inaccuracies in the 200 picks

case. It is more evident in the higher error rate and higher DOF case due to the solution performing the best in this situation.

Finally, in cases where it is difficult to estimate the probabilities, the information that must be known is only when events occur that may cause data to become inaccurate. This is demonstrated by the first set of analysis where we found no difference in the results when using a tally compared to using a static setting for the EP value.

## 7.1 Limitations

Whilst the proposed approach can help to avoid and find operational disruptions, there are various limitations. Firstly, the events that cause the inaccuracy need to be known in advance; although it is not necessary to know which particular instance of the event will cause an inaccuracy. Furthermore, the approach requires that each time the event occurs it is possible to automatically record that it has taken place, and also know which database records are the correct ones to tag. This may not be feasible or possible for some scenarios. Despite this reliance on the events, a benefit is that it is often inexpensive to simply record when an event occurs and keep track of this in a database. Contrast this with an RFID tracking solution that requires expensive hardware equipment to be installed. Another limitation is that the approach does not eradicate the disruptions, but continually avoids them. This is mitigated in some part with the ability to find the inaccuracies when the approach is configured this way; also the ability to avoid disruptions for as long as possible is certainly useful in various settings such as in times of peak load on a warehouse. However, a superior solution would be able to both detect and eradicate the inaccuracies. Also, the approach proposed requires that the events are independent and possess the memoryless property, which are limiting. The approach not being a proper metric of accuracy is also limiting as it is unknown whether the tag values can be used directly in a traditional data quality assessment. Another limitation relates to the simulation which did not measure disruption D4. The orders were chosen so that the warehouse can always satisfy the order because in prior executions of the simulation, it was found that many orders were not be picked when an order required more quantity to be available in the warehouse. In reality, many warehouses

have a separate area where items are stored in bulk (so the items are available, but just not easily accessible), and the picking locations (providing fast access to items) are replenished according to predictions of future orders. Hence, D4 is a rare issue compared to the others, and was therefore not measured in the simulation. A more general point about the choice of simulation as a methodological tool needs to also be raised. Even though simulations are often used to study warehousing processes like picking and replenishment, their limitations are known as their results can often be affected by the setting. We believe that this study has been designed in a way that allows a fair illustration of the benefits and limitations of our approach for a wide range of settings and cases and we note that analytical models could be used for further evaluation of this approach in future research.

## 8 Summary and Conclusion

This research has proposed and evaluated an alternative, and automatic, way of populating accuracy tags in information systems so that they can help to avoid operational disruptions and/or to find the inaccuracies that cause the disruptions. Both the avoid and find configurations of Potential Problem Data Tagging demonstrate an improvement over the normal case and increasingly so as the error rate and DOF increase. This brings the concept of data tagging closer to adoption in actual information systems because the approach reduces the level of data tag maintenance required due to it being highly automated. One advantage of the approach is that it is possible to switch between the avoiding and finding configurations dynamically, therefore, it is possible to better control when data with inaccuracies are encountered versus when the effects of poor data want to be avoided. This switching ability raises questions about when is the best time to perform a data quality detection/correction program, and deliberately “setting up” the data so that it is primed for a DQ detection/correction program. Finally, the choice of the warehousing decision-making task in this work illustrates how the experimental setting can be simplified in future data tagging research because 1. it is easy to measure whether the tags themselves are of benefit or not, 2. the decision-making task can be carried out automatically by the information system thus removing all “decision-maker

issues”, such as being overloaded and inconsistent, and 3. the tags do not need to be presented to the user, thus avoiding issues with representation and visualisation of tags (Shankaranarayanan and Zhu, 2016; Wang and Strong, 1996). Future work could consider extending these types of tags with values of utility in order that they can be changed and used for different contexts (Even and Shankaranarayanan, 2007) especially in Big Data environments (Merino et al., 2016), and this is important given that data is being repurposed for data analytics-related decision-making (Woodall, 2017) and performance reporting (Vallurupalli and Bose, 2018) at the same time as being used operationally.

## 9 Acknowledgements

We acknowledge YH Global Supply Chain Co., Ltd for supporting this research financially, and for providing access to their operations and expertise. We also thank Alena Puchkova for her help.

## 10 References

- Ardagna, D., Cappiello, C., Samá, W. and Vitali, M. (2018), “Context-Aware Data Quality Assessment for Big Data”, *Future Generation Computer Systems*, 89, pp. 548–562.
- Ballou, D.P. and Pazer, H.L. (1995), “Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff”, *Information Systems Research*, (6:1), pp. 51–72.
- Chan, H.-L., Choi, T.-M. and Hui, C.-L. (2012), “RFID versus Bar-Coding Systems: Transactions Errors in Health Care Apparel Inventory Control”, *Decision Support Systems*, (54:1), pp. 803–811.
- Chandola, V., Banerjee, A. and Kumar, V. (2009), “Anomaly Detection: A Survey”, *ACM Computing Surveys*, (41:3), pp. 1–58.
- Chengalur-Smith, I.N., Ballou, D.P. and Pazer, H.L. (1999), “The Impact of Data Quality Information on Decision Making: An Exploratory Analysis”, *IEEE Transactions on Knowledge and Data Engineering*, (11:6), pp. 853–864.
- DeHoratius, N. and Raman, A. (2008), “Inventory Record Inaccuracy: An Empirical Analysis”, *Management Science*, (54:4), pp. 627–641.

- English, L. (1999), *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, John Wiley & Sons.
- English, L. (2003), "Plain English About Information Quality: Defining and Measuring Accuracy", *DM Review*.
- English, L. (2009), *Information Quality Applied: Best Practices for Improving Business Information, Processes and Systems*, John Wiley & Sons.
- Even, A. and Shankaranarayanan, G. (2007), "Utility-driven assessment of data quality", *ACM SIGMIS Database*, (38:2), pp. 75–93.
- Fan, W. and Geerts, F. (2012), *Foundations of Data Quality Management*, Morgan & Claypool Publ, San Rafael, Calif.
- Faria, F. and Reis, V. (2015), "An Original Simulation Model to Improve the Order Picking Performance: Case Study of an Automated Warehouse", *Computational Logistics*, 9335, pp. 689–703.
- Fisher, C., Chengalur-Smith, I. and Ballou, D.P. (2003), "The Impact of Experience and Time on the Use of Data Quality Information in Decision Making", *Information Systems Research*, (14:2), p. 170–188.
- Fisher, C. and Kingma, B. (2001), "Criticality of Data Quality as Exemplified in Two Disasters", *Information & Management*, (39:2), pp. 109–116.
- Fisher, C., Lauria, E.J. and Matheus, C.C. (2009), "An Accuracy Metric: Percentages, Randomness, and Probabilities", *Journal of Data and Information Quality (JDIQ)*, (1:3), Article 16.
- Gelman, I.A. (2010), "Setting Priorities for Data Accuracy Improvements in Satisficing Decision-Making Scenarios: A Guiding Theory", *Decision Support Systems*, (48:4), pp. 507–520.
- Giannikas, V., Lu, W., Robertson, B. and McFarlane, D. (2017), "An Interventionist Strategy for Warehouse Order Picking: Evidence from Two Case Studies", *International Journal of Production Economics*, 189, pp. 63–76.
- Gilbert, G.N. and Troitzsch, K.G. (2005), *Simulation for the Social Scientist*, 2nd ed., Open University Press, Maidenhead, England ; New York, NY.

- Haegemans, T., Snoeck, M. and Lemahieu, W. (2016), "Towards a Precise Definition of Data Accuracy and a Justification for its Measure", *International Conference on Information Quality*, pp. 16–1.
- Han, J., Jiang, D. and Li, L. (2010), "Automatic Accuracy Assessment Via Hashing in Multiple-Source Environments", *Expert Systems with Applications*, (37:3), pp. 2609–2620.
- Heinrich, B. and Hristova, D. (2016), "A Quantitative Approach for Modelling the Influence of Currency of Information on Decision-Making Under Uncertainty", *Journal of Decision Systems*, (25:1), pp. 16–41.
- Heinrich, B., Hristova, D., Klier, M., Schiller, A. and Szubartowicz, A. (2018), "Requirements for Data Quality Metrics", *Journal of Data and Information Quality*, (9:2), pp. 1–32.
- Heinrich, B. and Klier, M. (2011), "Assessing Data Currency — A Probabilistic Approach", *Journal of Information Science*, (37:1), pp. 86–100.
- Heinrich, B. and Klier, M. (2015), "Metric-Based Data Quality Assessment — Developing and Evaluating a Probability-Based Currency Metric", *Decision Support Systems*, 72, pp. 82–96.
- Heinrich, B., Klier, M. and Kaiser, M. (2009), "A Procedure to Develop Metrics for Currency and its Application in CRM", *Journal of Data and Information Quality*, (1:1), pp. 1–28.
- Hipp, J., Müller, M., Hohendorff, J. and Naumann, F. (2007), "Rule-Based Measurement Of Data Quality In Nominal Data.", *International Conference on Information Quality (ICIQ)*, pp. 364–378.
- Hüner, K.M., Schierning, A., Otto, B. and Österle, H. (2011), "Product Data Quality in Supply Chains: The Case of Beiersdorf", *Electronic Markets*, (21:2), p. 141.
- Joglekar, N.R., Anderson, E.G. and Shankaranarayanan, G. (2013), "Accuracy of Aggregate Data in Distributed Project Settings: Model, Analysis and Implications", *Journal of Data and Information Quality*, (4:3), pp. 1–22.
- Kent, W. (1991), "The Breakdown of the Information Model in Multi-database Systems", *SIGMOD Rec.*, (20:4), pp. 10–15.
- Lee, H. and Özer, O. (2007), "Unlocking the Value of RFID", *Production and Operations Management*, (16:1), pp. 40–64.



- Lee, Y.W., Strong, D.M., Kahn, B.K. and Wang, R.Y. (2002), "AIMQ: A Methodology for Information Quality Assessment", *Information & Management*, (40:2), pp. 133–146.
- Lukyanenko, R., Parsons, J. and Wiersma, Y.F. (2014), "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content", *Information Systems Research*, (25:4), pp. 669–689.
- Marsden, J.R. and Pingry, D.E. (2018), "Numerical Data Quality in IS Research and the Implications for Replication", *Decision Support Systems*, 115, pp. A1–A7.
- Maydanchik, A. (2007), *Data Quality Assessment*, Technics Publications LLC.
- Merino, J., Caballero, I., Rivas, B., Serrano, M. and Piattini, M. (2016), "A Data Quality in Use model for Big Data", *Future Generation Computer Systems*, 63, pp. 123–130.
- Moges, H.-T., Vlasselaer, V.V., Lemahieu, W. and Baesens, B. (2016), "Determining the Use of Data Quality Metadata (DQM) for Decision Making Purposes and Its Impact on Decision Outcomes — An Exploratory Study", *Decision Support Systems*, 83, pp. 32–46.
- Montgomery, D.C. (2012), *Design and Analysis of Experiments*, 8th Edition International Student Version edition., John Wiley & Sons, Singapore.
- Motro, A. and Rakov, I. (1997), "Not All Answers are Equally Good: Estimating the Quality of Database Answers", *Flexible Query Answering Systems*, Springer US, Boston, MA, pp. 1–21.
- Nutt, P.C. (1998), "How Decision Makers Evaluate Alternatives and the Influence of Complexity", *Management Science*, (44:8), pp. 1148–1166.
- Olson, J.E. (2003), *Data Quality: The Accuracy Dimension*, Morgan Kaufmann, San Francisco.
- Parssian, A., Sarkar, S. and Jacob, V.S. (2004), "Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product", *Management Science*, (50:7), pp. 967–982.
- Parssian, A., Yeoh, W. and Ee, M.S. (2015), "Quality-Based SQL: Specifying Information Quality in Relational Database Queries", *Computer*, (48:9), pp. 69–74.

- Pedrielli, G., Vinsensius, A., Chew, E.P., Lee, L.H., Duri, A. and Haobin Li. (2016), "Hybrid Order Picking Strategies for Fashion E-Commerce Warehouse Systems", *Proceedings of the 2016 Winter Simulation Conference (WSC)*, pp. 2250–2261.
- Price, R. and Shanks, G. (2011), "The Impact of Data Quality Tags on Decision-Making Outcomes and Process", *Journal of the Association for Information Systems*, (12:4), p. 323.
- Redman, T.C. (1998), "The Impact of Poor Data Quality on the Typical Enterprise", *Communications of the ACM*, (41:2), pp. 79–82.
- Rekik, Y., Sahin, E. and Dallery, Y. (2008), "Analysis of the impact of the RFID technology on reducing product misplacement errors at retail stores", *International Journal of Production Economics*, (112:1), pp. 264–278.
- Sarac, A., Absi, N. and Dauzère-Pérès, S. (2010), "A Literature Review on the Impact of RFID Technologies on Supply Chain Management", *International Journal of Production Economics*, (128:1), pp. 77–95.
- Sessions, V. and Valtorta, M. (2009), "Towards a Method for Data Accuracy Assessment Utilizing a Bayesian Network Learning Algorithm", *Journal of Data and Information Quality (JDIQ)*, (1:3), p. Article 14.
- Shankaranarayan, G., Ziad, M. and Wang, R.Y. (2003), "Managing Data Quality in Dynamic Decision Environments: An Information Product Approach", *Journal of Database Management (JDM)*, (14:4), pp. 14–32.
- Shankaranarayanan, G. and Even, A. (2006), "The Metadata Enigma", *Communications of the ACM*, (49:2), pp. 88–94.
- Shankaranarayanan, G., Even, A. and Watts, S. (2006), "The Role of Process Metadata and Data Quality Perceptions in Decision-Making: An Empirical Framework and Investigation", *Journal of Information Technology Management*.
- Shankaranarayanan, G. and Zhu, B. (2016), "SPIDEQ - A Prototype for Decision Making with Quality Metadata", *Proceedings of the International Conference on Information Quality*, Spain, p. Paper 20.

- Suciu, D., Olteanu, D., Christopher, R. and Koch, C. (2011), *Probabilistic Databases*, 1st ed., Morgan & Claypool Publishers.
- Tu, Y.-J., Zhou, W. and Piramuthu, S. (2018), "A Novel Means to Address RFID Tag/Item Separation in Supply Chains", *Decision Support Systems*, 115, pp. 13–23.
- Vallurupalli, V. and Bose, I. (2018), "Business Intelligence for Performance Measurement: A Case Based Analysis", *Decision Support Systems*, 111, pp. 72–85.
- Wand, Y. and Wang, R. (1996), "Anchoring Data Quality Dimensions in Ontological Foundations", *Communications of the ACM*, (39:11), pp. 86–95.
- Wang, R., Reddy, M.P. and Kon, B. (1995), "Toward Quality Data: An Attribute-Based Approach", *Decision Support Systems*, (13:3–4), pp. 349–372.
- Wang, R.Y. and Strong, D.M. (1996), "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems*, (12:4), pp. 5–34.
- Watts, S., Shankaranarayanan, G. and Even, A. (2009), "Data Quality Assessment in Context: A Cognitive Perspective", *Decision Support Systems*, (48:1), pp. 202–211.
- Wechsler, A. and Even, A. (2012), "Using a Markov-Chain Model for Assessing Accuracy Degradation and Developing Data Maintenance Policies", *AMCIS 2012 Proceedings*.
- Woodall, P. (2017), "The Data Repurposing Challenge: New Pressures from Data Analytics", *Journal of Data and Information Quality*, (8:3–4), p. 11:1–11:4.
- Woodall, P., Oberhofer, M. and Borek, A. (2014), "A Classification of Data Quality Assessment and Improvement Methods", *International Journal of Information Quality*, (3:4), pp. 298–321.
- Yeganeh, N.K., Sadiq, S. and Sharaf, M.A. (2014), "A Framework for Data Quality Aware Query Systems", *Information Systems*, 46, pp. 24–44.
- Zak, Y. and Even, A. (2017), "Development and Evaluation of a Continuous-Time Markov Chain Model for Detecting and Handling Data Currency Declines", *Decision Support Systems*, 103, pp. 82–93.