

# Collecting and managing fuzzy data in statistical relational databases

Miroslav Hudec\* and Dušan Praženka

*Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia*

**Abstract.** Statistical institutes are focusing on variety of data sources from traditional surveys to big-data. Many of these data and concepts can be expressed as crisp values. But many other data cannot be expressed by precise values. In order to collect, store and manage the fuzziness in data we have adapted the fuzzy meta model as an extension of traditional relational database. Furthermore, experts' knowledge often contains vagueness and subjectivity. If we store this knowledge in a fuzzy database we can build knowledge management systems capable to cope with fuzziness. Statistical institutes cooperate in the data exchange. We have briefly discussed a simple way of extending the SDMX standard to cope with the fuzzy data in a way that does not influence exchanging precise values. Our research was focused on examining promising ways for managing fuzziness of real world because statistical institutes have been starting to analyze variety of promising data sources where not all data are always precise.

**Keywords:** Fuzziness of real world, fuzzy data in relational databases, knowledge management, data exchange by SDMX

## 1. Introduction

Nowadays, statistical institutes are launching traditional surveys but they are also looking for a variety of promising alternative data sources [1–3,6,25,31]. Many data are vague and imprecise (fuzzy) either by their nature, e.g. environmental data [33], or caused by non-ideal measuring [23]. The fuzziness is amplified when both types appear. Furthermore, many respondents' opinions either in surveys or on social networks are vague and imprecise. We could say that the values of attributes are not always known with the sufficient precision to justify the use of traditional relational databases to store them. Furthermore, calculations and results could be different when we do not neglect fuzziness in this data [11,15].

The same holds true for collecting, storing and managing information in knowledge management systems. Knowledge collected by experts often contains ele-

ments of uncertainty [26]. This information should be considered in knowledge base management systems and recommender systems [8,20,35]. It especially holds true for institutions where significant fluctuation of staff is recognized. Such institutions could be national and international statistical institutes.

Our objective is to create extension to existing relational database systems allowing efficiently storing and re-using fuzzy data related to official statistics. Full fuzzy database approaches [12] are sophisticated, but we see a lack of practical applications and tools [30]. On the other hand, relational databases are broadly used due to sound methodology, available CASE tools, vast literature and variety of relational database management systems. Hence, adding fuzziness into the relational databases is a promising way [16,17,30]. Working with fuzzy data has opened the problem of exchanging fuzzy data in a standardized way. We have also touched the problem of exchanging fuzzy data by the SDMX standard.

Even though the fuzziness is closely related to phenomena in social sciences and business, the mathematics of fuzzy logic is mainly applied in engineering systems and computer science [22]. Thanks to this trend,

---

\*Corresponding author: Miroslav Hudec, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemska cesta 1, 852 35 Bratislava, Slovakia. Tel.: +421 267295198; E-mail: miroslav.hudec@uba.sk.

engineering systems continue to be more and more sophisticated and powerful. We could reach the same in social sciences and official statistics, if we efficiently support them by fuzzy logic. Hence, we should not neglect these facts.

In order to meet the aforementioned aims, the paper is organized as follows. Section 2 explains some occurrences of fuzziness in the real world. Section 3 briefly explains fuzzy sets theory. Section 4 is devoted to managing fuzzy data in relational databases. Section 5 opens a question of possible adaptation of SDMX to exchange fuzzy information. Section 6 discusses way for minimizing impact of fuzziness on relational databases and SDMX. Section 7 speculates about building knowledge management systems supported by fuzzy logic. Finally, Section 8 concludes this paper.

## 2. Fuzziness in the real world

Statistical institutions are focusing on variety of data sources [32]: traditional surveys, administrative sources [5,10], alternative sources [6] (sensors [25], mobile positioning data [1,2], and social networks [31] among others) which measure development in the society and people's opinions about this development [3]. Many of these data and concepts can be expressed as sharp (crisp) numbers or clear expressions of natural language. But in the real world many other concepts cannot be expressed by precise values. Zadeh [36] expressed this fact as *do not know rationale*. It means that the values of attributes are not known with the sufficient precision to justify the use of traditional methods of numerical computing or to justify the use of traditional databases to store this data. In practice (ether in technical or social) we are not able to always precisely measure attribute's values of surveyed entities [23]. When data are vague, imprecise and ambiguous the uncertainty is not based on randomness, it cannot be presented as a crisp value. This type of uncertainty is called fuzziness [38] i.e. uncertainty in the semantic meaning of the analysed phenomena.

Many data are fuzzy either by their nature, caused by the tolerance level of instruments used for measuring, or as a result of respondents' estimation. For example "environmental data, quality of life data and measurements of continuous one-dimensional quantities cannot be adequately expressed by crisp numbers" [33]. A good example of the first type is the flooded level marked on a wall [16] illustrated in Fig. 1. Where ex-

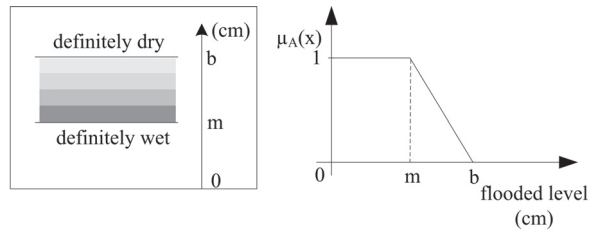


Fig. 1. Level of flood as a fuzzy set.

actly to measure this value in order to ensure that similar buildings will be similarly treated? By fuzzy data (right side in Fig. 1) we can avoid this problem. Values lower than  $m$  clearly states the full flood of a particular house, values greater than  $b$  clearly states that no flood was observed and values between  $m$  and  $b$  express intensity of flooding. Values closer to  $m$  from the right side have higher membership degree. We could easily imagine limitation of traditional data query or expert system (*if level is > 90 cm then...*). Value of 89 cm and 91 cm will be treated differently, even though their difference is questionable.

Further, we should keep in mind that the measurement made by a measuring instrument is usually approximate due to the tolerance interval [23]. It means that the precise value is somewhere in the (small) interval  $[a, b]$ , i.e.  $\mu(x) = 1$  for  $x \in [a, b]$ , where  $\mu(x)$  is function expressing belonging of element  $x$  to the interval.

People measure (guess) values often by estimation e.g. in questions related to the frequency of doing something during a year. For example, someone could declare that he/she used product approximately 90 days but for sure not less than 75 and not more than 100 days. Furthermore, respondents often estimate answers to the open-ended questions. In addition, it could be useful to allow skilled interviewers to remark relevance of answers or the credibility of surveyed person (intensity of record's belonging to a relational table).

Another example is expressing staff knowledge about solving tasks and how to find appropriate solutions for new tasks which are similar to some of the recorded ones [26]. Similarity of task and collected information is under consideration in the fuzzy recommender systems [13,20].

Hence, the following question arises: Why we should manage fuzzy data? We could just select one crisp value as a representative of measured value. It could cause the different treatment of similar entities [11,14,15].

This paper is not focused on searching for fuzzy data in variety of sources. The paper is rather dedicated to

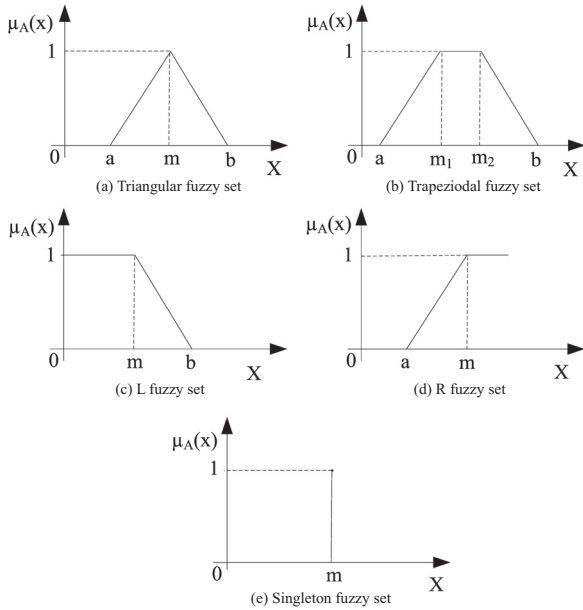


Fig. 2. Fuzzy sets.

managing fuzzy data in databases and data exchange. Our assumption is that relational databases (databases which are widely used) are able to manage traditional crisp data as well as fuzzy data in the same way.

In order to meet this goal, we have been worked on fuzzy relational database model. The next section explains theory of fuzzy set required for developing and using relational databases to manage fuzzy data.

### 3. Fuzzy sets and fuzzy logic in brief

The fuzzy set theory [37] provides a robust framework for systematically handling uncertainty based on fuzziness. In the fuzzy set theory belonging to a set is a matter of degree. A fuzzy set  $A$  over the universe of discourse  $X$  is defined by membership function  $\mu_A(x)$  that matches each element of the universe of discourse  $X$  with its membership degree to the set  $A$  in the following way:

$$\mu_A(x) : X \rightarrow [0, 1] \quad (1)$$

Each fuzzy set (usually expressed by linguistic term) is defined by one membership function. Concepts like *medium value* or *value close to  $m$* , where  $m$  is a real number are expressed by triangular or trapezoidal fuzzy sets (Figs 2a and 2b respectively). *Small value* (decreasing) is expressed by L fuzzy set (Fig. 2c) whereas *high value* (increasing) is explained by linear

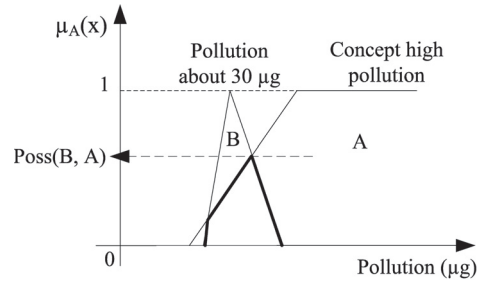


Fig. 3. Illustration of possibility function.

gamma (R) fuzzy set (Fig. 2d). Finally, we should consider sharp value of fuzzy data as singleton fuzzy set (Fig. 2e) and not as a crisp value. The reason for this decision is explained later on.

During the work with fuzzy data we need to compare fuzzy value with a fuzzy concept. For this purpose we can use possibility and necessity measures or the comparison operator [11]. The last one produces fuzzy set as a result. Therefore, we should add the defuzzification step. Possibility that the fuzzy value  $B$  belongs to a fuzzy concept  $A$  is defined as [11] (Fig. 3):

$$Poss(B, A) = \sup_{x \in X} [t(A(x), B(x))] \quad (2)$$

where  $X$  is a universe of disclosure and  $t$  is a t-norm [19]. In practice, usually minimum t-norm is used:

$$t_{\min} = \min(A(x), B(x)) \quad (3)$$

The possibility measure gets value from the  $[0, 1]$  interval. Figure 3 shows possibility that fuzzy value *pollution about 30 µg* belongs to the concept *high pollution*.

Value approximately 30 could be expressed as crisp value 30, interval or triangular fuzzy set. In case of crisp value, similar entities are not always adequately treated [15]. The difference between interval and fuzzy number is shown in Fig. 4. Values near edges of the interval belong to the interval in a same way as values situated in the middle of the interval. Therefore, we should carefully create intervals. In a fuzzy way values near the edges of fuzzy set belong to the set with significantly lower degree than values in the middle. It means that values of  $a$  and  $b$  (Fig. 2a) do not influence solution in the same way as edges ( $I_L, I_H$ , Fig. 4) of intervals. Hence, entities having values  $p_1$  and  $p_2$  are indistinguishable in the interval (both belong to the interval), whereas these entities are clearly distinguishable in fuzzy set (different membership degrees).

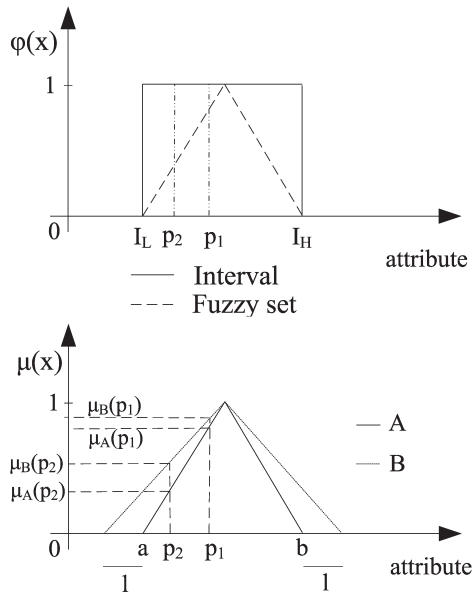


Fig. 4. Interval and fuzzy data.

If we have condition like *attribute A is near 30 and attribute b is near 50 and attribute c is medium* then membership degrees in all fuzzy sets ensures natural way of sorting results. The *and* connective is expressed by t-norms [19]. Therefore, we do not need to calculate distance to the middle of interval to properly sort all records.

#### 4. Fuzzy data in relational databases

Collected data are usually stored in databases. Broadly used databases are relational ones. Therefore, this section is focused on adjusting relational databases to cope with fuzzy data.

##### 4.1. Preliminaries

The existing papers in this field are either focused on theoretical aspects clearly stating that future design of databases should not neglect the fact that many data are fuzzy [33] or on complex fuzzy database models and solutions.

Concerning fuzzy databases, two main approaches: possibility based models and similarity based models were developed. In the former approach all the knowledge about values of an attribute  $A$  (defined in the domain  $D$ ) of an entity  $x$  (record or row in a database) can be represented by the possibility distribution  $\pi_A(x)$  on the domain  $D$  [12,27]. In the later approach a fuzzy

Table 1  
Basic model

id	attribute 1	attribute 2	...	m
1	Small	551		0.85
2	Medium	296		1
3	Small	535		0.75
4	High	125		0.25

relation is defined as a subset of an attribute domain. Several similar values of an attribute explain particular entity [12,24]. The GEFRED model [21] is a synthesis of above explained models. It is based on generalized fuzzy domain of an attribute and generalized fuzzy relationships between database tables.

These models are sophisticated in order to cope with fuzziness in attributes as well as in relationships. However, they lack broad real applications due to the complicated structure and the lack of available tools like for traditional relational databases [30]. In the real world we can find many situations where relationships between entities are crisp, but only values of several attributes are fuzzy. In this case the adaptation of traditional relational databases is a rational option.

##### 4.2. Fuzziness in traditional relational databases

The significant limitation for managing fuzzy data is the first normal form. This form declares that in the database cell only atomic value can be stored [7]. For crisp numeric values it is the optimal way. On the other hand, fuzzy values contain several parameters and therefore, they should be stored in a different way to satisfy the requirements of the first normal form and other integrity rules.

###### 4.2.1. Basic model

This is the simplest way for including fuzziness. A relational table is extended with the column expressing intensity of belonging to the relation for each record. An illustrative example is in Table 1.

The benefit is the simple structure. Just one attribute (column  $m$  in Table 1) is added. The domain of this attribute is the  $(0, 1]$  interval. The interval is open from the left side, because adding rows which do not belong to a relation does not have any sense. For instance, this way allows skilled interviewers to remark relevance of answers or the credibility of surveyed person.

The main drawback is the lack of more deep information related to fuzziness. We are not sure whether is uncertainty related to the whole record or to the particular attribute(s). For example, if we do not have documentation from the cadastre, then the attribute *age of*

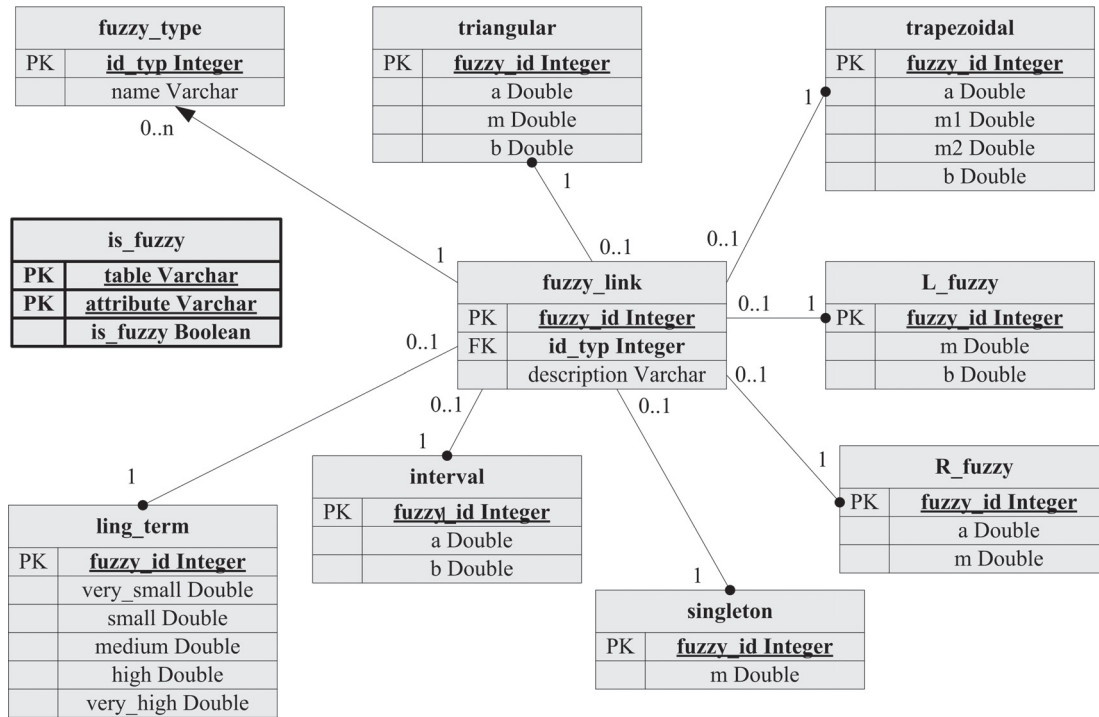


Fig. 5. Fuzzy meta model adjusted from [29].

Table 2  
Flexible query on basic model

id	attribute 1	attribute 2	$\mu_{attr.2}$	m	$\mu_{query}$
1	Small	551	1	0.85	0.85
3	Small	535	0.70	0.75	0.70

building can be vague (e.g. estimated according to the specific architectural elements).

Data retrieving is a relatively simple task. In case of crisp query (e.g. attribute 1 = “small”) value  $m$  show us the intensity of belonging to a query for each selected record. In this case belonging to a query cannot be higher than belonging to a relation. We can also use column  $m$  in a query condition (e.g. select rows significantly belonging to a relation).

Flexible SQL query conditions [17,24] can be straightforwardly added to this model. Let us have flexible condition *where attribute 2 is high* and let the term *high* is defined as fuzzy set with parameters  $a = 500$  and  $b = 550$  (Fig. 2d). The intensity to satisfying query is calculated by t-norm function (3) between satisfying the query condition and belonging to a relation in the way shown in Table 2.

#### 4.2.2. Fuzzy meta model

Each fuzzy data is represented by membership function (usually several parameters, except for the single-

ton fuzzy set – Fig. 2e). If these parameters are stored in a way that requirements for normality and integrity are satisfied, then we can manage fuzzy data in a relational database. The solution for this idea is the fuzzy meta model which manages fuzzy data and links to the relational tables of real entities [16,29,30]. Real entities and their crisp attributes are stored in relational tables in the usual way. Therefore, this way does not affect attributes containing only precise values. Collecting, storing, updating and retrieving this attributes remains the same as without the fuzzy meta model.

The fuzzy meta model is shown in Fig. 5 [29]. Two main tables are *is\_fuzzy* stating which attributes in each table are fuzzy and *fuzzy\_link* which is related to the attribute in relational model of real entities. Table *fuzzy\_type* allows us to express type of fuzzy set for each record (triangular, trapezoidal, linguistic term, etc.). Each fuzzy value is represented by membership function. For example, the table *trapezoidal* has the following attributes (fuzzy\_id, a, m1, m2, b) in order to manage storing trapezoidal fuzzy data (Fig. 2b). Opinions could be expressed by linguistic terms e.g. *very low*, *low*, *medium*, *high* and *very high*. When respondents cannot explicitly state which linguistic term is the most suitable they tend to express answer as e.g. *more medium than low* which can be expressed by

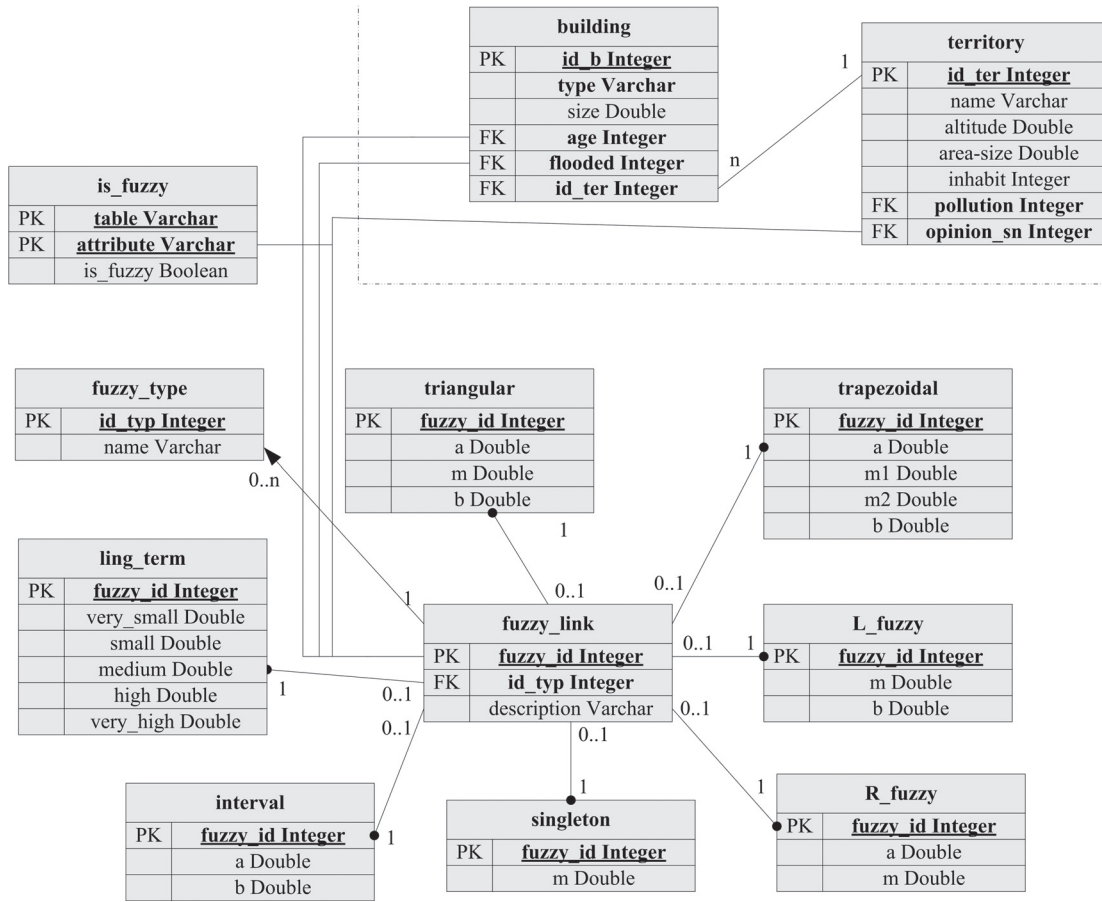


Fig. 6. Model for storing fuzzy data of territorial units in relational database.

membership degrees [14]. If we would like to store uncertainty about flooded level (Fig. 1), then the choice is L fuzzy set (Fig. 2c) where parameters  $m$  and  $b$  have the same meaning as in Fig. 1.

The idea of fuzzy meta model [29] is demonstrated on the statistics related to the buildings and territorial units because this statistics contains traditional attributes and moreover, variety of big-data can be related to territorial units such as opinions on social networks, climatic conditions, pollutions, data from positioning mobile devices, energy consumption. Logical model of this, for illustrative purpose, statistical database is shown in Fig. 6.

In the table *territory* attributes *pollution* and *opinion* on social networks (*opinion\_sn*) do not store real values but foreign keys to the respective tables where real values are stored as fuzzy. The same holds for the attributes *age* and *flooded* in the table *building*. Other attributes are stored as crisp values in the relational tables of entities. In this way relational database is ca-

pable to store crisp and fuzzy data and meet integrity rules.

#### 4.2.3. Storing new records and retrieving existing ones

Storing new records in the database is explained by the following pseudo code [16]:

```

check whether attribute appears in table
is_fuzzy;
if attribute is fuzzy then
    generate ID for foreign key;
    check acceptable fuzzy sets for the attribute
    and offer them in a combo box;
    according to selected type of fuzzy set
    provide input boxes for parameters;
    verify input values and store them in
    relevant tables;
else
    allow storing crisp value for attribute;
end if

```

building

id	type	size	age	flooded	id_ter
1	t 1	150	1101	2101	1
2	t 2	285	1102	2102	1

territory

id_ter	name	altitude	area_size	inhabit	pollution	opinion_sn
1	mun 1	125	13200	5385	3101	4101
2	mun 2	583	12905	4297	3102	4102

is\_fuzzy

table	attribute	is_fuzzy
building	type	0
building	size	0
building	age	1
building	flooded	1
territory	name	0
territory	altitude	0
territory	area_size	0
territory	inhabit	0
territory	pollution	1
territory	opinion_sn	1

fuzzy\_link

fuzzy_id	id_type
1101	triangular
1102	singleton
2101	l
2102	singleton
3101	triangular
3102	triangular
4101	ling_term
4102	ling_term

triangular

fuzzy_id	a	m	b
1101	140	150	160
3101	35	40	45
3102	90	100	105

l\_fuzzy

fuzzy_id	a	b
2101	90	105

singleton

fuzzy_id	m
1102	42
2102	0

ling\_term

fuzzy_id	very_small	small	medium	high	very_high
4101	0	0	0.3	0.7	0
4102	0.85	0.15	0	0	0

Fig. 7. Data in fuzzy relational database.

The size of stored data is affected by number of fuzzy attributes (columns) and types of fuzzy sets. In the extreme situation all  $n$  numeric attributes (excluding primary keys) for all  $m$  entities are fuzzy of trapezoidal form. Therefore, in this case the size of database is  $9(m \cdot n)$  [16]. Anyway, user should carefully decide which attributes should be stored as fuzzy and in which form. Keeping in mind this deduction, we can say that the size of fuzzy relational database can be significantly lower than this extreme situation. The relational instance of a fuzzy relational database is shown

in Fig. 7.

From the tables shown in Fig. 7 we can easily reconstruct all fuzzy as well as crisp data. Building 1 belongs to the territory 1, size is 150 m<sup>2</sup>, age is about 150 years but for sure not lower than 140 and not higher than 160, building was flooded in a way that under 90 cm wall was definitely wet and above 105 cm was definitely dry. Pollution in territory 1 is about 40 mg of measured pollutant but for sure not lower than 35 mg and not higher than 45 mg. Opinion about this territory on social networks is rather positive.

Because data are stored in a relational database, classical SQL queries, fuzzy queries as well as more sophisticated approaches such as summaries from the data [14] can be used. In processes of retrieving by queries we have to calculate intensity of belonging to the query when condition and/or queried attribute are fuzzy. In flexible queries membership degrees or possibility and necessity functions are used, among others [11]. These functions allows us to calculate how fuzzy data e.g. *pollution about 30 mg* belongs to the concept e.g. *high pollution*. It is useful in querying, disseminating and data processing.

#### 4.2.4. Including fuzziness into the existing databases

Adding fuzziness into a relational database could be straightforwardly realized anytime. In case of fuzzifying database which already contains data firstly, we have to decide which attributes will be collected as fuzzy. For example, after several years of using traditional database we have recorded many complaints about measuring values and therefore, managing financial compensations related to floods and pollution. If we have the following rule managing knowledge related to financial support:

*if flooded level is > 100 cm and pollution > 50 mg  
then provide 5 000*

then the limitation of sharp rule and sharp data is clear. Measured value of 98.5 cm and 49.5 mg cause no financial support whereas values of 101 cm and 50.5 mg cause full support. To avoid this drawback we can create the following rules:

*if flooded level is between 80 and 100 cm and pollution is between 35 and 50 mg then provide 2 500  
if flooded level is > 100 cm and pollution > 50 mg then provide 5 000  
etc.*

but we have only mitigated strong jump into the less strong one. In order to improve our knowledge of this problem we could decide to record fuzziness in attributes and use flexible rules in a way:

*if flooded level is high and pollution is high then provide high financial support*

Intensities of belonging of fuzzy data into fuzzy concepts influence the intensity of financial support by one simple rule. An example of managing fuzzy rules on municipal statistics database can be found in [18].

When attributes *flooded* and *pollution* are converted into the fuzzy ones, the previously collected data are

migrated to fuzzy data of singleton type. A unique value is generated to replace data in the entity's table with the foreign key. These keys point to the table *singleton* where we can find real value migrated from the relational table of entities. At this point the database is ready to manage fuzzy data for new data collections.

## 5. Fuzzy data in standards for data exchange

Another interesting question is the exchange of uncertain data and information among statistical organizations if needed. SDMX standard is widely used for exchange of aggregated data [4,9]. The experimental tool [28] has shown that SDMX can be also used in managing micro data. If institutes would like to exchange fuzzy data, then we could slightly improve SDMX standard.

If we drew analogy between model explained in Figs 5 and 6 and the SDMX standard [9], then it implies that we could add attribute informing about fuzziness on the measurement level. Definitions of dimensions are not affected by fuzziness. Information related to fuzziness could be linked on the observation level explaining whether value is precise or fuzzy (e.g. OBS\_TYPE = "P" or OBS\_TYPE = "F").

Therefore, on the measurement level we should add tag which informs users whether the measured variable contains fuzzy information. Further, for each observation we should add tags explaining type of fuzzy information (fuzzy set) and parameters of fuzzy set. Currently, these tags does not exist but they construction could be possible if institutions create consensus for using fuzzy data. An illustrative example of possible tags is as follows:

```
<sts_c: series .....Metadata.....>
<sts_c: Obs TIME_PERIOD = "2012"
OBS_VALUE = "583" OBS_STATUS = "A"
OBS_TYPE = "P">
<sts_c: Obs TIME_PERIOD = "2012"
OBS_VALUE = "1101" OBS_STATUS =
"A" OBS_TYPE = "F">
<sts_c: Obs_Detail ID = "1101"
FUZZY_TYPE = "Trian"
OBS_VALUE_A = "35"
OBS_VALUE_M = "40"
OBS_VALUE_B = "45">
</sts_c: series>
```

The OBS\_TYPE tag says whether value is precise (P) or fuzzy (F). We could arrange that if this tag does not exist, then the value is by default precise. It ensures



full compatibility with the current SDMX structure. If OBS\_TYPE is F, then the OBS\_VALUE tag is not a value but is a key for the next sublevel where fuzzy value is explained. Analogously holds for storing fuzzy data into relational database (Figs 6, 7). In our illustrative SDMX example the second row contains fuzzy data expressed as triangular fuzzy sets by parameters  $a = 35$ ,  $m = 40$  and  $b = 45$ . In the same way textual data such as opinions and their intensities (Fig. 7) can be modelled. Comparing this code with Figs 5 and 7 we can conclude that conversion from relational table to SDMX and opposite can be realized.

## 6. Discussion

In Sections 4 and 5 we have explained possibilities for storing and exchanging fuzzy data. Suggested approach does not force statistical institutions to work with fuzzy data. In the future more and more data will be interesting for official statistics. We could expect that not all data expressing some aspects of the real world can be forced to be crisp without loss of relevant information.

We would like to emphasize that a fuzzy meta model can be added to the existing database anytime, either for new attributes or existing ones. In case of existing attributes, previously collected data can be transformed by simple procedures. It does not affect attributes which remain to be collected as sharp values. Fuzzy meta model can be realized in any relational database management system.

Concerning new SDMX tags, the prerequisite is mutual agreement in the SDMX consortium to build these tags in a way that managing precise values in SDMX format is not affected. If no fuzzy data is in exchange we can just omit tags explaining fuzzy data. Currently, this is not a hot topic. But in the future, we could expect large variety of data from different sources. Moreover, issues explained on illustrative attributes (Section 2) could appear for many other attributes.

## 7. Knowledge systems in relational databases supported by fuzzy logic

Fuzzy data and fuzzy rules could significantly support knowledge base management systems. Knowledge about entities and managing rules can be expressed by natural language. Let us consider database rows as cases and columns expressing values (terms)

of relevant attributes. If new case for which we do not know the solution appears, we can search for most similar cases in a database [8,11,26]. In a traditional database we can detect whether two rows contain the same values. But when values are vague, subjective and imprecise (as is usual with knowledge) then finding most similar rows is not an easy task by traditional tools. Fortunately, fuzzy logic provides variety of approaches like fuzzy functional dependencies, conformance between attributes, similarity between their values, flexible implications and variety of aggregations. Some of these functions were applied in new approach for calculating fuzzy functional dependencies [34]. The future task will be experimenting on the data and knowledge base related to the public administration and statistics.

Another direction is building flexible recommender systems [13,20,35]. These approaches for official statistics might be promising but further research is definitely required. For the collecting experts' opinions and vague information we could use aforementioned fuzzy meta model. However, if the requirements for managing more complex types of fuzziness appear, then approaches such as similarity based models or GEFRED should be applied.

## 8. Conclusion

It should not be neglected that many data of statistical interest are fuzzy. We cannot always force them to be expressed as sharp values without loss of relevant information. Furthermore, crisp values cannot be always ideally measured or estimated. Therefore, we need an efficient way for storing and re-using these values in variety of analyses. Considering that, our work was focused on managing fuzzy data in traditional relational databases, because these database systems are widely used.

The task has been solved by adding tables (fuzzy meta model) and creating relevant validation rules. Each fuzzy data is represented by a set of parameters of respective membership function. Number of fuzzy attributes and types of fuzzy sets affects the size of database. Anyway, users could reasonably decide which attributes should be managed as fuzzy and which should remain to be crisp.

Knowledge base management systems are based on experts' knowledge and data. The former is often expressed by vague terms. The later cannot be always sharp value. Hence, fuzzy databases could be a suit-

able support tool for knowledge management and recommendation systems.

Nowadays, statistical institutes are focusing interest on variety of sources from traditional surveys to big-data. In addition, statistical institutes are highly cooperating in exchanging data and metadata. For exchanging data and metadata the SDMX standard has been created. We have discussed a way for extending SDMX to cope with fuzzy data in a way that exchange of precise data is not influenced.

To summarize, we have not discussed a final software tool but rather promising concepts for coping with the fuzziness of real word.

## References

- [1] R. Ahas, M. Tiru, E. Saluveer and C. Demunter, Mobile telephones and mobile positioning data as source for statistics: Estonian experiences. Proc New Techniques and Technologies for Statistics (NTTS). Brussels; 2011.
- [2] L. Altin, M. Tiru, E. Saluveer and A. Puura, Using passive mobile positioning data in tourism and population statistics. Proc New Techniques and Technologies for Statistics (NTTS). Brussels; 2015.
- [3] S. Balbi, M. Hudec, J. Juriová, M. Klúčik, A. Stawinoga and N. Triunfo, Report on Analysis of Existing Practices in the Data Collection Field, Deliverable 5.2, Blue-Ets Project (FP7); 2013.
- [4] G. Becker and M. Bruschi, An SDMX-based unified data catalogue (UDC). Proc UNECE/OECD/Eurostat Meeting on Management of Statistical Information Systems. Dajeon; 2010.
- [5] L. Costanzo, G. Di Bella, E. Hargreaves, H.J. Pereira and S. Rodrigues, An overview of the use of administrative data for business statistics in Europe. Proc 58th World Statistical Congress. Dublin; 2011.
- [6] P. Daas, M. Roos, C. De Blois, R. Hoekstra, O. Ten Bosch and Y. Ma, New data sources for statistics: experiences at Statistics Netherlands. The Hague/Heerlen: Statistics Netherlands; 2011.
- [7] C.J. Date, Date on databases: Writings 2000–2006. New York: Apress; 2006.
- [8] G. De Tre, T. Matthe, P. KordJamshidi and M. Demoor, On the use of case based reasoning techniques in flexible querying. Proc 18th International Workshop on Database and Expert Systems Applications. Regensburg; 2007.
- [9] Eurostat, Introduction to SDMX, Student Book, Eurostat Directorate B. Luxembourg: Eurostat; 2010.
- [10] J. Dunne and J. Hayes, Realising the statistical potential of administrative data. Proc Seminar on New Frontiers for Statistical Data Collection. Geneva; 2012.
- [11] J. Galindo, editor. Handbook of research on fuzzy information processing in databases. Hershey: Information Science Reference; 2008.
- [12] J. Galindo, A. Urrutia and M. Piattini, Fuzzy databases – Modeling, design and implementation. Hershey: Idea Group Publishing; 2006.
- [13] E. Herrera-Viedma, C. Porcel, A. López-Herrera and S. Alonso, A fuzzy linguistic recommender system to advice research resources in university digital libraries. Studies in fuzziness and soft computing, volume 220. Springer; 2008. pp. 567–585.
- [14] M. Hudec, Storing and analysing fuzzy data from surveys by relational databases and fuzzy logic approaches. Proc XXV-th IEEE International Conference on Information, Communication and Automation Technologies. Sarajevo; 2015.
- [15] M. Hudec, Managing fuzziness of real world in business informatics. Proc Strategic Management and Support by Information Systems. Uherské Hradiště; 2015.
- [16] M. Hudec, Fuzzy data in traditional relational databases. Proc 12th Symposium on Neural Network Applications in Electrical Engineering. Belgrade; 2014.
- [17] M. Hudec, Fuzzy database queries in official statistics: Perspective of using linguistic terms in query conditions, *Stat J IAOS* 29(4) (2013), 315–323.
- [18] M. Hudec and M. Vujošević, A fuzzy system for municipalities classification, *Central European J. Operations Res* 18(2) (2010), 171–180.
- [19] E.P. Klement, R. Mesiar and E. Pap, Triangular norms. Dordrecht: Kluwer Academic Publishers; 2000.
- [20] J. Lu, D. Wu, M. Mao, W. Wang and G. Zhang, Recommender system application developments: A survey, *Decision Support Syst* 74 (2015), 12–32.
- [21] J.M. Medina, O. Pons and M.A. Villa, GEFRED: A generalized model of fuzzy relational databases, *Inf Sci* 76(1–2) (1994), 87–109.
- [22] A. Meyer and H.J. Zimmermann, Applications of fuzzy technology in business intelligence, *Int J Comput Communicat Control* VI(3) (2011), 428–441.
- [23] F. Pavese, Why should correction values be better known than the measurand true value? *J. of Physics, Conf. series* 459; 2013.
- [24] F. Petry, Fuzzy databases – Principles and applications. Boston: Kluwer; 1996.
- [25] M. Puts, P. Daas and M. Tennekes, High frequency road sensor data for official statistics. Proc New Techniques and Technologies for Statistics (NTTS). Brussels; 2015.
- [26] L. Portinale and A. Verrua, Exploiting Fuzzy-SQL in case-based reasoning. Proc 14th International Florida Artificial Intelligence Research Society Conference (FLAIRS). Key West; 2001.
- [27] H. Prade and C. Testemale, Fuzzy relational databases: Representational issues and reduction using similarity measures, *J. American Society Inf. Sci.* 38(2) (1987), 118–126.
- [28] D. Praženka and P. Boško, Combining technical standards for statistical business processes from end-to-end. Proc New Techniques and Technologies for Statistics (NTTS). Brussels; 2011.
- [29] S. Škrbić, Using fuzzy logic in relational databases. Ph.D. Dissertation, University of Novi Sad; 2008.
- [30] S. Škrbić, M. Racković and A. Takači, Towards the methodology for development of fuzzy relational database applications, *Comput. Sci. Inf. Syst.* 8(1) (2011), 27–40.
- [31] V. Torres van Grinsven and G. Snijkers, Sentiments and perceptions of business respondents on social media: an exploratory analysis, *J. Official Statist.* 31(2) (2015), 283–304.
- [32] S. Vale, Using administrative and secondary sources for official statistics - A handbook of principles and practices. Geneva: UNECE; 2011.
- [33] R. Viertl, Fuzzy data and information systems. Proc WSEAS International Conference on Systems, Corfu; 2011.
- [34] M. Vucetic, M. Hudec and M. Vujošević, A new method for computing fuzzy functional dependencies in relational

- database systems, *Expert Syst. Appl.* **40**(7) (2013), 2738–2745.
- [35] R.R. Yager, Fuzzy logic methods in recommender systems, *Fuzzy Sets Syst.* **136**(2) (2003), 133–149.
- [36] L.A. Zadeh, From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions, in: *Computing with words*, P. Wang, ed., New York: John Wiley & Sons; 2001. pp. 35–68.
- [37] L.A. Zadeh, Fuzzy sets, *Inf. Control* **8**(3) (1965), 338–353.
- [38] H.J. Zimmermann, Fuzzy set theory and its applications, London: Kluwer Academic Publishers; 2001.