

Frank, A. U. "Procedure to Select the Best Dataset for a Task." Paper presented at the Third International Conference on Geographic Information Science GIScience 2004, Maryland, October 20-23 2004.

## Procedure to Select the Best Dataset for a Task

Andrew U. Frank\*, Eva Grum\*, and Bérengère Vasseur\*\*

\* Institute for Geoinformation and Cartography, Technical University of Vienna,  
Gusshausstrasse 27-29, 1040 Vienna, Austria  
frank@geoinfo.tuwien.ac.at, grum@geoinfo.tuwien.ac.at

\*\* Université de Provence, Laboratoire des Sciences de l'Information et des Systems  
LSIS,  
Rue Joliot-Curie, 13453 Marseille cedex 13, France  
berengere.vasseur@lsis.org

**Abstract.** This paper models the decision process when selecting among different datasets the one most suitable for a task. It shows how metadata describing the quality of the dataset and descriptions of the task are used to make this decision. A simple comparison of task requirements and available data quality is supplemented with general, common-sense knowledge about effects of errors, lack of precision in the data and the dilution of quality over time. It consists of two steps: first, compute the data quality considering the time elapsed since the data collection; and second, assess the utility of the available data for the decision. A practical example of an assessment of the suitability of two datasets for two different tasks is computed and leads to the intuitively expected result.

### 1 Introduction

The selection of a dataset or a map from a choice of several potentially useful ones seems to be nothing else than a numerical comparison between the requirements of the task with the available data quality listed in the metadata description. It is often done intuitively, but a computational model for this seemingly simple task is missing. A computational model is necessary for search engines to use the metadata for automatic ranking of available sources with respect to an intended use. Unless search engines use metadata, especially data about geographic data quality, the investment in metadata production does not yield the expected benefits.

The collection of metadata is often advocated, for example in the EU project INSPIRE, but metadata is not always available and sometimes in a grotesquely unsuitable form (Hunter and Goodchild 1995). One may conclude that they are not often used (UNE 2001). Is the lack of practical interest in metadata due to the absence of use? Do we really know how to use data quality descriptions? In this paper, we give a computational model for the use of metadata describing the quality of geographic data to assess the usability or "fitness for use" (Chrisman 1984). We assume a situation where more than one dataset from different sources could be useful for a task. The task may be navigation in town, where the decisions are to turn left or right at a street corner or it may be the decision about the location of a facility. A

dataset is contributing to a task if it improves the decision. Valuable data lead to an execution that uses fewer resources or produces a better result (Krek 2002).

The rational, programmable method proposed here leads to the selection of the datasets that provide most improvements in the decision for the task. The approach is novel in several aspects:

- it is a quantitative assessment,
- it breaks user groups into individual tasks and decisions,
- it is based on qualitative models of processes (Kuipers 1994),
- it separates temporal currency from other descriptions of data quality, and
- it leads to a programmable selection using existing metadata.

We describe here a rational method for comparing the utility of different datasets with respect to an intended use. It is based on a detailed assessment of the data quality for each theme available in the data sets following the usual model of data quality description for geographic data (Moellering 1987; Chrisman 1988; Morrison 1988; Goodchild and Gopal 1990). A decision process connects the data with the task. In this process the quality of the data is transformed into the quality of the task. A model of this decision process contains all the information necessary to assess the utility of a dataset to improve the task. We use this “common sense” knowledge when we make an intuitive decision about which data source to use for a decision.

The paper demonstrates the method using a simplified example: two data sets are available and assessed for use by a tourist and firefighters. The data sets are patterned after the multi-purpose digital map of Vienna and a digital tourist map (Wilmerdorf 1992; Wien 2004). The resulting recommendation agrees with our intuitive choice and are formalized, ready to be integrated into a search engine for geographic data.

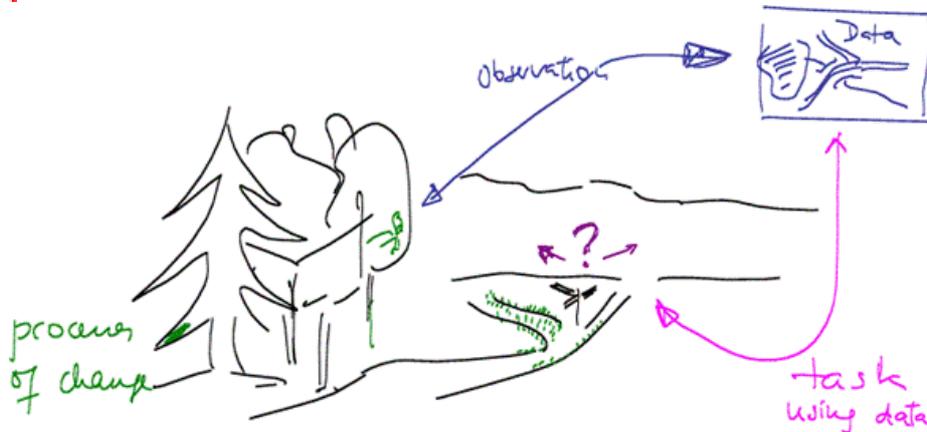


Fig. 1: Closed loop semantics (Frank 2003): the observation process is linked to the task using the data (section 3). Changes occur in the world between data collection and execution of task (section 4) and task using data for decisions (section 5)

The paper is structured as follows (fig. 1): The second section of the paper describes the terminology used and how we model the question in the abstract. It covers the user task and its requirements; what is data quality and how to describe the usability of a map. It introduces the processes involved, namely data collection,

changes in the world, and decision. The third section assesses the available data. We assume that geometric precision, attribute accuracy and temporal currency are the most dominant components of data quality. The data quality description therefore concentrates on these aspects. It is an abstract description of the observation and data collection processes. Section 4 then describes how the data quality assessment is updated, considering the usual changes in the world. Section 5 then assesses the usability of a single data set for a single decision. Section 6 summarizes the result and gives a comprehensive description of the process that leads to the selection of a dataset for a decision and applies it quantitatively to an example. The seventh section introduces the notion of a user group and shows why a decision among several datasets is usually not made for individual tasks, but for user groups. Section 8 concentrates on the issue of clutter: a data set that contains irrelevant data is less suitable than one that has exactly what is required. The concluding section summarizes the result and points out that the same method can be used as a guideline to design new usable geographic information products.

## 2 Processes

In this section we explain how we conceptualize the situation and introduce the terms we use subsequently.

A user needs to make a decision in a task and considers the acquisition of a dataset (or a conventional map). The example task in this paper is navigation in a city, that is, the decision about turning left or right at a street corner, extensively studied by Krek (2002). Other spatial decisions have different requirements for the data necessary, but they can be dealt with the same logic.

The user has the choice of several data sets that are potentially beneficial for her task. They have *utility* for a task if they lead to making better decisions and a better execution of the task. The data sets contain data describing different aspects of reality, which we will call *themes*. Examples are: street name, the position of an ATM or the number of the bus line serving a bus stop. We use the term data *quality* to describe the correspondence between an object in reality and its representation in the data set. Data quality will be differentiated for several *aspects*, such as precision and completeness.

The description of the quality of the data available to the user is a description of the processes that were used for the collection; the quality of the data is directly linked to these processes (Timpf, et al. 1996). This description of data quality is independent of the task; the provider of the data can only describe the data from this perspective.

Originally, the data collection process is linked to the decision process for which the data are collected, which is abstracted in the description of the transformation of data quality to utility of the decision. In this *closed loop* from the real world to data collection, to data use for a decision, to a task executed in the real world (fig. 1), semantics of data quality measures can be defined (Frank 2003). Data are usually collected long before they are used in a decision; reality is changing in the intervening interval. These changes must be modeled as a decrease of the data quality (section 4).

The models of processes used here are second order; the first order process is the decision process itself. Therefore, only rough estimates of how these (second order)

processes of data quality transformation work are sufficient. The model of the influence of data quality on decision quality leads to a rational assessment of usability of a dataset, even in the absence of detailed knowledge.

Theme	Multi purpose map			City map		
	when collected	complete	precision (m)	when collected	complete	precision (m)
Bus line				2001	90%	0,5
Church				1998	99%	0,1
Fire hydrant	2001	99%	0,05			
Building number	2003	99%		2000	90%	
Points of interest				1999	90%	0,5
Stations for public transport				2002	95%	1
Street name	2001	95%		1999	95%	
Street network	2002	98%	0,1	2003	95%	0,2
Tram line				2001	90%	0,2
Walls between and inside buildings	2003	99%	0,05			

Table 1: Example of data quality description (fictitious values)

### 3 Description of the Data Quality of a Data Set

This section describes how we assess the quality of the data sets when it is collected. It follows the literature on metadata and data quality descriptions (Moellering 1987; Chrisman 1988; Morrison 1988; Goodchild and Gopal 1990). The quality of a dataset results from the processes used for data collection. Data collected with the same processes have similar quality; it is therefore reasonable to give data quality indications for whole data collections or large parts that were collected in the same way.

The measurement and observation procedures determine the *precision* of the data, which are typically described by the standard deviation of random errors – assuming no gross errors and no systematic bias. The data collection methods used determine the *completeness* of a theme. For brevity, completeness is used as a summary assessment for omissions and commissions; precision of data not measured on a continuous scale is lumped together with completeness. Last but not least, the *date* of the data collection indicates when the correspondence between reality and the data were assessed by precision and completeness.

The two datasets used here as examples for typical content in publicly available maps useful for in-city navigation are the multi-purpose map of the city of Vienna and the city map of Vienna (Wien 2004). Table 1 lists some of the themes in these

datasets and fictitious quality assessments for them. The data quality assessment values are for demonstration purposes only and do not reflect the data quality found in the products distributed by the Magistrate Vienna.

## 4 Dilution of Precision and Completeness over Time

The metadata that come with a dataset describe the data quality at the time of collection or last major update. With time the world changes and the quality is reduced (figure 2). For a decision about the utility of a dataset, the quality today is of interest. This section shows how to update the given precision and completeness with estimated rates of dilution in quality. It addresses the paradoxes that the data do not change with time, but nevertheless their quality dilutes (data do not age with time like good wine!).

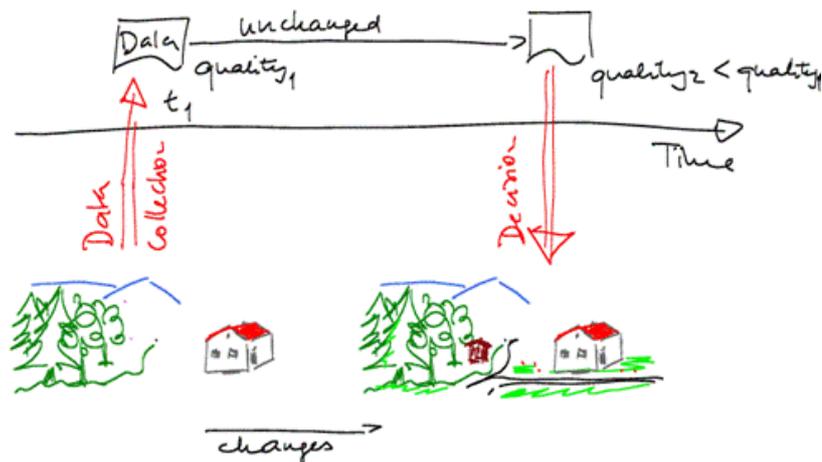


Fig. 2: The world changes, but data remains fixed

Figure 2 gives a more understandable explanation of the previously published statement that precision, completeness and up-to-datedness are *not* orthogonal (Frank 1987; Frank 1998). The utility of a dataset of high precision, but observed some time ago, and a not-so-precise dataset observed more recently is perhaps the same.

### 4.1 Precision

The precision of the location of objects is diluted by random movements. Such movements are more important for natural themes, like forest boundary, streams etc., than for the manmade objects in a city (Burrough 1996). The precision is reduced proportionally to the time intervened:

$$p_i = p_0 - r_p * dt_i$$

where  $dt_i = t_i - t_0$

$t_0$  the time of data collection

$p_0$  the precision at that time

$r_p$  the rate of change in the precision

Similar to location, the precision of other values can be transformed to a current date considering the rate of dilution in the precision of this data element. Some descriptive values, for example prices, change gradually over time by the inflation rate, which is a systematic, first-order effect and can be taken into account when transforming the data. The dilution in precision is due to the differential change in prices, which reduces the precision with which we can know the current price given the price several years ago, second order effect. Estimation of quantitative values is not crucial; we leave it for future work to study the influence of errors on these estimates.

Data quality description for a data set	$r_c$		$r_p$			Multi purpose map			City map of Vienna		
	change	move	complete	correct	precision	complete	correct	precision			
Building number	2%		97%	97%	0	98%	77%	0			
Bus line	5%	0.01	0%	100%	0	105%	100%	0.47			
Church	1%	0.01	0%	100%	0	105%	100%	0.04			
Fire hydrant	4%	0.01	87%	87%	0.02	0%	100%	0			
Points of interest	4%	0.01	0%	100%	0	110%	100%	0.45			
Stops of public transport	5%	0.01	0%	100%	0	105%	100%	0.98			
Street name	1%		92%	92%	0.08	95%	90%	0			
Street network	5%	0.01	88%	88%	0	100%	100%	0.19			
Tram line	4%	0.01	0%	100%	0	102%	100%	0.17			
Walls between buildings	3%	0.01	96%	96%	0.04	0%	100%	0			

Table 2: Updated (diluted) data quality

## 4.2 Completeness

The completeness is diluted by the rate of new objects appearing and old ones disappearing (a different factor could be used to transform omissions and commissions separately).

$$c_i = c_0 - r_c * dt_i \quad \text{where } c_0 \text{ the completeness at the time of data collection}$$

Table 2 gives the assumed values for rate of change of precision and completeness.

## 5 Description of User Task

One cannot determine the suitability of data for a task just by studying the data quality description alone, but has to consider the task and the decision that should be made with the information from the dataset and how it is influenced by the quality of the data. A compact description of the user's decision situation and how it is influenced by the available quality of the data is necessary. The description of the task given here is reduced to the aspects relevant for the selection of a data set considering data quality alone and is only one component of a comprehensive description of a task. This is the goal of this section.

Krek has shown how the quality of spatial information affects the quality of a decision (Krek and Frank 2000; Krek 2002); for a given decision, it is possible to identify which themes influence the decision. Only the data quality of the themes that influence the decision affects the quality of the decision.

Krek computed the value of a dataset as the economic contribution it makes to improve the decision and thus the outcome of the task, comparing with the decision made without information available. Here, we want to assess the utility of a dataset and compare it with the perfect dataset, that is, the dataset that contains no error, which has by definition utility 1. Any other dataset has a lesser utility, expressed as a percentage. The unusable dataset has utility 0.

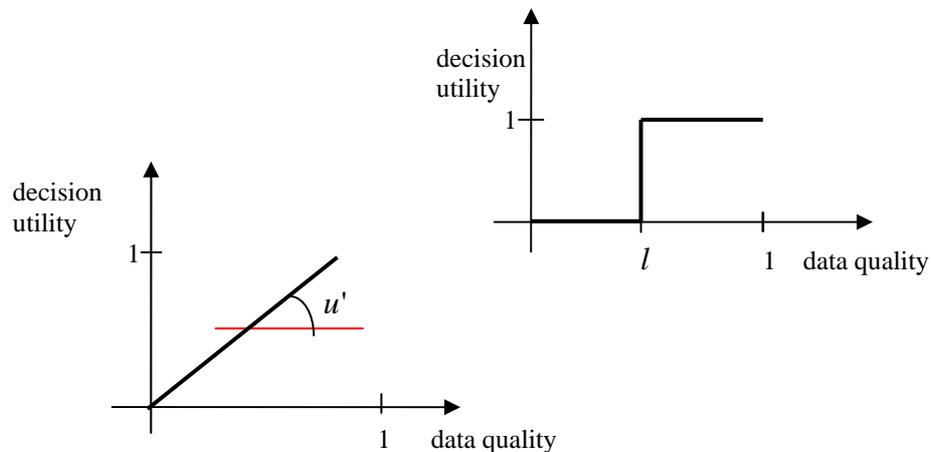


Fig. 3: The two types of influence of data quality: gradual and threshold

### 5.1 Influence of Single Decision: Precision and Correctness of one Theme

The influence of the quality of the dataset on the decision depends on the dataset and the decision. Two typical cases can be differentiated (figure 3):

- **gradual influence:** “the higher the quality, the better the decision;” it is described with the rate  $u'$  with which the utility increases with an increase in the data quality.
- **threshold:** if the quality is above a certain level, the decision is the same and the utility 1; if the quality is lower, then the decision is the same as if there were no data, that is, the utility is 0; it is characterized by the threshold value  $l$ .

These two influences are sometimes combined: up to a threshold the quality does not influence the decision, and from then it increases with the quality proportionally (Fig. 4). Note that precision is usually expressed such that lower values mean higher quality!

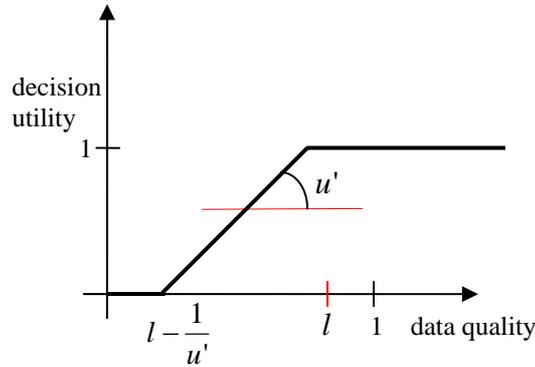


Fig. 4: The influence model with a threshold value and an influence rate

$$u_{p_i} = \begin{cases} 0 & \text{for } p_i < l - \frac{1}{u'} \\ 1 & \text{for } p_i > l \\ 1 - (l - p_i) \cdot u' & \text{otherwise} \end{cases}$$

where  $u_{p_i}$  is the resulting utility from precision,  $p_i$  the available data precision,  $l$  the threshold and  $u'$  the rate of increase.

### 5.1 Influence on Many Decisions: Completeness of one Theme

The formula above computes the influences of the data quality for a single decision where the data are available. If the data are not available, the utility is 0. The average utility given for an incomplete dataset is the result of multiplying  $u$  with the completeness factor.

$$u = u_p * c \quad \text{where } u \text{ is average utility, } c \text{ is the completeness}$$

There may be cases where the lack of data is interpreted as negative fact, invoking the closed world assumption (Reiter 1984). Then the decision may be worse than in the absence of any information—a negative utility! This is not further considered here, because it is related to the discussion of omission and commissions, which is left for future studies.

### 5.3 Multiple Themes Used in one Decision

Decisions may require several themes as inputs. Two different situations can be differentiated:

**Single decision:** Each theme contributes individually and independently to the decision. The utility of the decision is the weighted average of the contributions of the themes. The weighted average is used because the utility is normalized to the interval [0..1].

**Related decisions:** One or several themes are enabling. If the data are not present, data from other themes cannot be used. This is in decision making sometimes called a K.O. criterion and related to Maslow’s pyramid, where requirements on a higher level are only considered if the lower level requirements are fulfilled (Jahn 2004). If a theme is crucial and the utility of the theme for the decision is less than a low threshold (say 20%), then the total utility of the dataset for this decision is 0.

### 5.4 Summary and Example

Every decision related to a task is characterized by four parameters:

- Threshold  $l$  for data quality to count as available.
- Rate of increase  $u'$  in utility with data quality.
- Is it a K.O. criterion?
- Relative weight compared to other themes.

For three example decisions, the themes used and the parameters are shown in table 3:

Task		Threshold $l$	$u'$	K.O.	Weight
<b>Navigation to a street</b>	Street name			1	1
	Street network	5	0.1	1	1
<b>Find building on street</b>	Street name				1
	Building number	2	0.2	1	1
<b>Public transportation use</b>	Stop of public transportation	10	0.1	1	1
	Bus lines	20	0.1		1
	Tram lines	20	0.1	1	1

Table 3: Characterization of tasks

## 6 Calculate Utility for a Specific Decision

A specific decision requires only few inputs themes. For these the utility is computed and summed up. Table 4 shows the computation for the task *navigation to a street*. For the two themes *street name* and *street network* the contribution of the theme to this decision is computed; it uses the updated data quality from table 2 and the task description from table 3 and computes from right to left the usability due to precision and correctness and then multiplies it with the completeness value. The contributions for the two themes are then averaged and yield the total utility.

Navigation to a street	Task description			Multi-purpose map			
	Theme	Threshold	$u'$	K.O.	complete	correctness	precision
Street name				1	92%	95%	0
contributions					87%	95%	1
Street network		5	0.1	1	88%	100%	0.08
contributions					88%	100%	1
<b>Utility total</b>					<b>88%</b>		

Table 4: Utility of themes for a task

With the same formulae, the utility for other tasks and datasets can be computed as in table 5.

Task	Multi-purpose map	City map of Vienna
navigation to a street	88%	83%
find building on street	90%	70%
public transportation use	0%	79%
fire response planning	92%	0%
find interesting places	30%	58%

Table 5: Utility for different tasks

The multi-purpose map of Vienna is assumed to be more up-to-date, which results in a higher utility for a task in navigation to a street and finding a building on a street; the higher geometric precision does not influence the result at all.

## 7 User Groups

In the previous sections, decisions for the suitability of a dataset are made for a single task and a single decision. This is, for example, the case when a dataset is acquired for a single administrative decision.

In practice a decision to acquire a dataset or a map is often made in order to cope with a complex mixture of future decisions including contingencies, like where is the next pharmacy, or where is the next police station. The cost of acquiring a dataset and the cost of learning how to read and use it is considerable. A multi-purpose set, which one can use in many situations and for many tasks—even if not completely optimized for my present task—is preferable than to learning each time how to use the specially constructed data set for this task.

Traditionally, maps were produced for special user groups. There are city maps for tourists and for locals; special maps are prepared for emergency services. Topographic maps are produced to serve a very wide audience. Recently some more specialized maps have appeared, such as maps with bicycle paths, or maps with bigger letters for the elderly.

A user group is characterized by a weighted set of task and the related decisions. The utility of a dataset is assessed as the weighted mean of the utility of the dataset for each of the decisions. For example, the user group “tourists” relates to the tasks:

- identify interesting objects, and find locations where they are clustered,
- finding an object to visit on the map, and
- use the map for navigation and identification of the object.

Firefighters have different needs: they are interested in buildings and space to drive to a building as close as possible, independent of traffic regulations, the location of fire hydrants and fire resistant walls inside buildings. Some of the tasks in a user group can be declared as K.O. criteria and the utility of a dataset is zero if this criterion is not satisfied to some reasonable degree.

Table 6 gives the utility of the two maps for different tasks of importance to the user groups and gives the weights for each task and user group. With these weights, the multi-purpose map is most suitable for the fire-fighters and the city map suits the tourist best (table 7). If K.O. criteria were assigned, the same result would be

	Utility		Weight	
	Multi purpose map	City map of Vienna	Tourist	Firefighter
Navigation to a street	88%	83%	3	4
Find building on street	90%	70%	2	4
Public transportation use	0%	79%	3	0
Fire response planning	92%	0%	0	3
Find interesting places	30%	58%	4	0

Table 6: Tasks for two user groups, the utility of two datasets for these tasks and weights two user groups assign to these tasks.

	tourist	firefighter
Multi purpose map	47%	<b>89%</b>
City map of Vienna	<b>72%</b>	56%

Table 7: The overall utility of two datasets and two user groups

## 8 Clutter Reduces Usability

If two datasets contain all the data required, but one offers additional data that are not required for any decision of this user group, the two datasets are numerically assessed the same utility, but intuitively they are not equally usable. We prefer the dataset with less clutter. The additional effort to filter out the unwanted data is a cost for the user and may lead to errors in decisions.

To reduce usability due to clutter we can reduce the usability by a factor proportional to the amount of data required, compared with the amount of data presented. The clutter factor is less than 1 and is multiplied with the usability measure introduced above. To calculate clutter ratio, the number of data elements per theme must be available in the metadata.

$$\text{clutter ratio} = \frac{\text{data used}}{\text{data presented}}$$

For the example data sets and user groups, we find that the clutter factor for the tourist using the multi-purpose map is much lower (worse) than for the city map (table 8). This reflects the intuitively higher usability of the city map for tourists.

Map elements		Multi-purpose map	clutter factor	City map	clutter factor
offered		1919.84		739.33	
required	firefighter	1796.74	<b>0.93</b>	724.65	<b>0.98</b>
	tourist	801.32	<b>0.42</b>	733.83	<b>0.99</b>

Table 8: Usability reduced for clutter

## 9 Conclusion

This method to calculate the utility of a data set for a task emerged from a method to describe suitability of a dataset for a task in a matrix, where on one axis the available data themes were listed and on the other axis the requirements for the task. It was refined by separating the assessment of the dataset into one table and the required data quality into another table (Grum and Vasseur 2004). This reduced the effort to prepare data quality and task descriptions from  $(n \times m)$  to  $(n + m)$ , where  $n$  is the number of task and  $m$  is the number of themes.

The new method is based on a model of the processes involved and a very generalized qualitative assessment of these processes (Kuipers 1994). Only rough estimates of the quantities involved are sufficient to the metadata into an assessment of the utility of the dataset for a task or user group.

The method can be used—as described—to select the most useful dataset for a specific task. It can be used equally well when considering the design of an

information product (Krek and Frank 1999). It leads to a multi-step procedure for a design of an information product:

- Identify a user group.
- List the tasks the users have to solve and identify the decisions included.
- For each decision, consider information inputs that can improve the decision.

The research reported here creates the framework for precise and operational definitions for data quality measures. Open questions relate to the treatment of omissions and commissions, with respect to objects and attributes recorded for these objects. Another long standing question is how to describe the quality of data not measured on a continuous scale

Understanding the decisions related to a task is quite limited. Two Ph.D. candidates are working on a detailed analysis—at the level of detail demonstrated by Hutchins (Hutchins 1995)—of two tasks, namely in town movement, using public transportation, and orienteering in open landscape.

The method is based on quantitative estimates of the influence of data quality on decisions; it should be investigated how sensitive the decision between two datasets is to variations in these estimates.

## Acknowledgements

Many of my colleagues and students have helped me advance to this point. Of importance where the many discussion in the REVIGIS project, led by Robert Jeansoulin. Particular thanks go to Werner Kuhn and Max Egenhofer for valuable discussion over the years. The comments from three reviewers were very helpful to improve this final version of the paper.

This work was funded as part of the REVIGIS project #IST-1999-14189. We also thank the magistrate Wien for providing us with datasets and descriptions and apologiz for the liberty we have taken with assessing the quality in this demonstration.

## References

1. Burrough, P. A. (1996). Natural Objects with Indeterminate Boundaries. *Geographic Objects with Indeterminate Boundaries*. P. A. Burrough and A. U. Frank (eds.). London, Taylor & Francis: 3-28.
2. Chrisman, N. R. (1984). "The role of quality information in the long-term functioning of a geographic information system." *Cartographica* 21(3 and 4): 79-87.
3. Frank, A. U. (1987). *Overlay Processing in Spatial Information Systems*. Eighth International Symposium on Computer-Assisted Cartography (AUTO-CARTO 8), Baltimore, MD: 16-31.
4. Frank, A. U. (1998). Metamodels for Data Quality Description. *Data quality in Geographic Information - From Error to Uncertainty*. R. Jeansoulin and M. Goodchild (eds.). Paris, Editions Hermès: 15-29.

5. Frank, A. U. (2003). Ontology for spatio-temporal databases. *Spatiotemporal Databases: The Chorochronos Approach*. M. Koubarakis, T. Sellis and et alia Berlin, Springer-Verlag LNCS 2520: 9-78.
6. Goodchild, M. F. and S. Gopal (1990). *The Accuracy of Spatial Databases*. London, Taylor & Francis.
7. Grum, E. and B. Vasseur (2004). How to select the best dataset for a task? *International Symposium on Spatial Data Quality 04*. A. U. Frank. Vienna: 197-206.
8. Hunter, G. H. and M. F. Goodchild (1995). "Dealing with Error in a Spatial Database: A Simple Case Study." *Photogrammetric Engineering & Remote Sensing* 61(5): 529-537.
9. Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA, The MIT Press.
10. INSPIRE *IN*frastructure for *SP*atial *IN*foRmation in *EU*rope [http]. 2003 [cited 14.07.2004]. Available from <http://www.ec-gis.org/inspire/>.
11. Jahn, M. (2004). User Needs in a Maslow Schemata. *International Symposium on Spatial Data Quality 04*. A. U. Frank. Vienna: 169-182.
12. Krek, A. (2002). An agent-based model for quantifying the economic value of geographic information. Ph.D. *Department of Geoinformation*. Vienna, Austria. Technical University Vienna.
13. Krek, A. and A. U. Frank (1999). *Optimization of Quality of Geoinformation products*. SIRC 99, Dunedin, New Zealand: 151-159.
14. Krek, A. and A. U. Frank (2000). "The Economic Value of Geo Information." *Geo-Information-Systeme* 13(3): 10-12.
15. Kuipers, B. (1994). *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. Cambridge, MA, The MIT Press.
16. Moellering, H. (1987). A draft proposed standard for digital cartographic data, National Committee for Digital Cartographic Standards.
17. Morrison, J. (1988). "The proposed standard for digital cartographic data." *American Cartographer* 15(1): 9-140.
18. Reiter, R. (1984). Towards a logical reconstruction of relational database theory. *On Conceptual Modelling, Perspectives from Artificial Intelligence, Databases, and Programming Languages*. M. L. Brodie, M. Mylopoulos and L. Schmidt (eds.). New York, Springer Verlag: 191-233.
19. Timpf, S., M. Raubal, et alia (1996). *Experiences with Metadata*. 7th International Symposium on Spatial Data Handling, SDH'96, Delft, The Netherlands (August 12-16, 1996), IGU: 12B.31-43.
20. UNE. *The Use of Metadata in University Libraries and Campuses* [http]. 2001 [cited 28.06.04 2004]. Available from <http://www.caul.edu.au/surveys/Metadata.doc>.
21. Wien, Magistrat (2004). Maps of the city of Vienna.
22. Wilmersdorf, E. (1992). The Implementation of A Multidisciplinary GIS-Network for An Urban Geocommunication System. Proceedings of EGIS '92, Munich, EGIS Foundation: 1408-1416.