JUDGING AND VISUALISING THE QUALITY OF SPATIO-TEMPORAL DATA ON THE KAKAMEGA-NANDI FOREST AREA IN WEST KENYA

K. Huth, N. Mitchell, G. Schaab

Karlsruhe University of Applied Sciences, Faculty of Geomatics, Moltkestraße 30, D-76133 Karlsruhe, Germany
corresponding author: gertrud.schaab@hs-karlsruhe.de

ABSTRACT:

This paper discusses the judgement of data quality for a collection of 132 geodatasets and evaluates different methods of visualising the results for presentation alongside the datasets themselves in a visualisation tool. The data has been accumulated within the BIOTA East Africa subproject E02 for research into the longer-term forest cover change for the Kakamega-Nandi forest complex in west Kenya and requires quality assessment to enable future critical use of the geodatabase. The data sources comprise satellite imagery, aerial photography, topographic maps and drafts, forestry maps, map sketches, thematic maps, and datasets derived from fieldwork including oral histories. The database includes image, raster and vector datasets, all of which are judged individually by one person according to the six data quality parameters of lineage, positional accuracy, attribute accuracy, logical consistency, completeness, and temporal information. Their selection is based on a literature review. Four of the judgements make use of a 1 to 5 ranking scale. For five of the six parameters additional information is given. The statistical assessment reveals distinct patterns. For example, the older maps generally have a poorer positional accuracy, and forestry maps are often both incomplete and have a poor date reliability. A second literature review helps to select ten different cartographic methods of visualising the quality parameters in diagrammatic form. These are reviewed against such criteria as universal suitability, memorability, and the potential for easy combination with text. The methods are then assessed to select the best medium in which to illustrate the six quality criteria in a complex diagram for each dataset. With memorability, clarity and ease of use foremost in mind, a traffic light system of visualisisation is selected as the best option for five quality parameters while a slider is chosen to present the completeness parameter. With the addition of the crispness variable the diagram is able to reflect the state of our knowledge of temporal information. The paper thus introduces a geodata quality visualisation concept for a real life example of a comprehensive database including geodatasets of a temporal dimension.

1. INTRODUCTION

Modelled predictions for 2100 reveal that the largest impact on biodiversity is expected to be due to land use/cover change (LUCC), this being especially true for the tropics (Sala & Chapin 2000; Chapin III et al., 2000). Within the BIOTA East Africa research framework, funded since 2001 by the German Federal Ministry of Education and Research (BMBF) (Köhler, 2004; http://www.biota-africa.org), the influence of fragmentation and anthropogenic use on the biodiversity of East African rainforests is investigated. With fifteen subprojects at present, BIOTA-East is following an integrated and interdisciplinary approach. Research is related to the vegetation structure, ecological interactions, certain animal groups (emphasizing invertebrates) and, since 2005, also to socio-economic issues in order to work towards a sustainable use of biodiversity (Schaab et al., 2005). The focal site is Kakamega Forest in west Kenya with Mabira and Budongo Forests in Uganda also selected for comparative purposes with research mainly based on 1km^2 biodiversity observatories (BDOs).

The BIOTA East Africa subproject E02 at Karlsruhe University of Applied Sciences supports this biodiversity research with geographic information system (GIS) and remote sensing activities aiming at an extrapolation of the field-based findings in space and time (Schaab et al., 2004). Here, E02 considers the analysis of longer-term forest cover changes in the three East African rainforests as one of its major research tasks. Data sources range from satellite imagery and historical aerial photography via old topographic maps, official governmental records and forestry maps to oral testimonies by the local population, with place names giving evidence for much earlier

forest extents (Mitchell & Schaab, 2006; Mitchell & Schaab, subm.). The analysis of such information will lead to a detailed picture of the forest use history for the different forests. The Nandi Forests are also included here as the development of the land use/cover time series brought them to light as having once been connected to Kakamega Forest (Mitchell, Lung & Schaab, 2006). The data gathered reflect approximately the last 100 years, coinciding with the start of commercial-scale exploitation of forests in East Africa.

The geodata processing so far is most advanced for the Kakamega-Nandi forest complex. A total of 132 data layers are directly visually compared as well as jointly analysed via their spatial reference by means of a GIS. The reliability of the
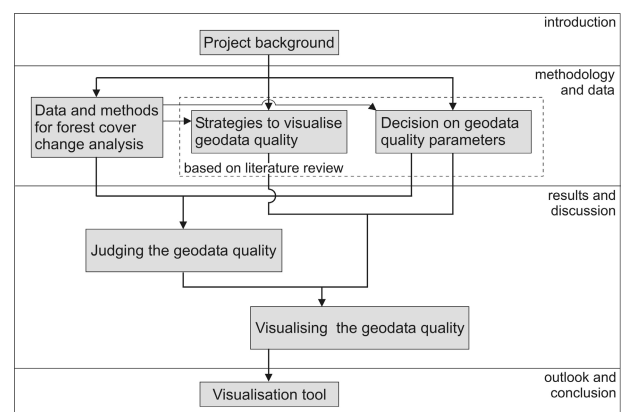

Figure 1. Structure of the paper.

spatio-temporal information must be accounted for and differences in geodata quality must be assessed by the scientist in order to draw correct conclusions. The quality of the time-stage dependent spatial data also needs to be documented and preferably visually cognisable for simply describing the underlying data layers and for the presentation of conclusions. This has lead to the concept of a visualisation tool including a feature that, since this data would be available, enables a consideration and visualisation of geodata quality.

This paper (Figure 1) starts with a description of the data sources and methodology applied for analysing the longer-term forest cover change. It reviews and discusses strategies for the visualisation of geodata quality. And finally, based on a crisp literature review we will conclude which data quality parameters are considered of importance for the purpose of our work. Next, our system of judging these parameters for every geodataset is introduced and a statistical summary is given on the judgements of all the Kakamega-Nandi datasets. Subsequently evolved designs for diagrams exposing the quality aspects are presented and discussed. An assessment of the alternatives leads to the agreed version for illustrating the group of geodata quality parameters. The outlook will stretch to the implementation in the visualisation tool.

2. ANALYSIS OF LONGER-TERM FOREST COVER CHANGE

In order to analyse the long-term change in vegetation cover datasets were sought to cover a 100-year time period. This necessitated the acquisition of spatial data in many different formats and consequently their integration within a GIS (see Table 1). Thus, the most recent period is covered by Landsat satellite imagery (MSS, TM and ETM+) which was purchased to allow analysis of the forest cover from the present day back to 1972 in 8 approximately five-year time-steps. A supervised multispectral classification was performed for each of the time-steps of Landsat satellite imagery. This process distinguished 12 land cover classes, six of which are forest formations and two of which are grassland (Lung, 2004; Lung & Schaab, 2004).

Historical aerial photography was also acquired to extend the time series back to 1965/67 and 1948/52, although the latter represents only 65% coverage of the forests. The photographs were scanned, orthorectified and mosaiced and from this land cover classes were distinguished by visual interpretation and were digitised on-screen. Vegetation classes were assigned in keeping with those derived from the satellite imagery (Mitchell et al., 2006).

Extending the series beyond remote sensing has required topographic maps from archives in Kenya and libraries in Britain. The search produced 15 topographic maps or map drafts which pertain to the Kakamega-Nandi forest area across the period 1896 to 1970 (ranging in scale from 1: 50,000 to 1: 1.5 million). In some cases the original map was acquired but for the most part they exist as photocopies or scans and occasionally as amateur digital photography. These were georeferenced for inclusion in the GIS and their relevant features, such as forest cover, were also digitised on-screen.

Forestry maps and logging records were painstakingly located in the forest offices of the Kenyan Forest Department. These forestry maps relate to the period 1933 to 1995 and can show both areas of logging and planting of trees. Some of these are printed maps while others are here termed forestry sketches. Logging concessions, for instance, were often sketched onto other maps by a forester using a coloured crayon, while in other cases they exist as tracings, hand-drawn sketches, or even as written descriptions of the boundary with reference to local landmarks. Logging records have been extracted from forestry archives and are incomplete but are linked to the concession maps. All the maps have been scanned, georeferenced and digitised to include relevant depictions or textual annotations of vegetation cover.

Other maps labelled here as sketch maps are present as hand drawings by e.g. anthropologists and date from between 1924 and 1949. There are 13 thematic maps which range from 1899 to 2000 and depict various themes from land use cover and population density to tribal locations. Fieldwork represents some of the most recent datasets and includes oral histories which were obtained by means of 69 semi-structured interviews with old people living adjacent to the forests. The interviews' locations have been established while a summary table of the main issues investigated is linked to the point layer. Other datasets derived from fieldwork but still in preparation include placename evidence and ground truth information.

To date, a total of 132 datasets are stored in the GIS covering the past 111 years (Table 1) but it should be emphasised that many of the geodatasets have incomplete coverage of the Kakamega-Nandi forests. Attention should also be drawn to the fact that several vector datasets can be derived from the

Data source	Combined date range*	Scale range	Image / Raster	Vector
Satellite imagery	1972 - 2003 (32 years)	30 to 60 m	14 / 8	-
Aerial photography	1948 - 1991 (44 years)	1:25k to 1:55k	3 / -	2
Topographic maps	1900 - 1970 (71 years)	1:50k to 1:1mill	7 / -	10
Topographic drafts	1896 - 1970 (75 years)	1:62.5k to 1:1.5mill	8 / -	9
Forestry maps	1937 - 1995 (59 years)	1:10k to 1:50k	8 / -	10
Forestry sketch maps	1933 - 1977 (45 years)	1:10k to 1:62.5k	9 / -	18
Sketch maps	1924 - 1949 (26 years)	1:62.5k to 1:300k	3 / -	9
Thematic maps	1899 - 2000 (102 years)	1:25k to 1:2.5mill	3 / -	10
Fieldwork	2002 - 2006 (5 years)	1:50k	-	1
Totals	**1896 – 2006 (111 years)**	**1:10k to 1:2.5mill**	**63**	**69**

* Date range refers to the 'date on map/geodataset'

Table 1. Geodatasets available for the analysis of longer-term forest cover change in the Kakamega-Nandi area.

digitising of the features of a single scanned map. In the case of satellite imagery, scenes from different seasons have in some cases been combined to represent a single timestep in order to enhance the classification of land cover types.

3. REVIEWING AND DISCUSSING STRATEGIES TO VISUALISE GEODATA QUALITY

With such diverse geodatasets at hand, a literature review was performed in order to gain ideas for the presentation of their quality. Different sources on the topic geodata quality served to find out which aspects of geodata quality can be treated and how others have visualised geodata quality in their projects.

Many examples in different contexts were found (for the complete compilation see Huth, 2007). In classifying the existing visualisations we first consider what kind of information is presented. There are those which show only one criterion, for example positional accuracy while others just refer to the overall data quality. For a summary see van der Wel et al. (1994). In particular interactive information systems can present information on several parameters, sometimes even with different levels of detail. With such features the software Quality Information Management Model (QIMM; Devillers et al., 2005) is already approximating a GIS. Others allow the user to choose between different, sophisticated visualisation alternatives (QIMM or RVIS, see below). Such interactive tools are not only rather complex in development but also in their correct use.

For simple visualisation of a single quality feature or the overall geodata quality, methods known from traditional cartography are quite common. These are e.g. the reliability diagram or the indicatrix by Tissot (van der Wel et al., 1994). Adapted to the electronical presentation of geodata are the methods using sound or blinking but also the well-known dot animation by Fisher (1994). An example of a more extensive tool is the software Reliablility Visualization System (RVIS) developed by MacEachren et al. (1996). Here the user can choose between different visualisation alternatives for the same dataset focussing on spatial, temporal and attribute quality aspects. Even more complex is the aforementioned QIMM by Devillers et al. (2005). Within this tool it is possible to show quality information for different levels of detail and six distinct quality parameters. Their display can be realised in the main map or as a quality dashboard next to the map.

Many of the examined methods are only suitable for particular types of geodata. RVIS for example is only designed for one special dataset and is thus restricted in its application. It is therefore difficult to transfer these particular methods to other geodatasets.

Geodata quality information can be either visualised within the map or map display or it can be placed independently from the map face. For the first option it is necessary to have differentiated quality information available for different areas in the map, e.g. applying transparency for depicting uncertain areas by MacEachren's variable 'clarity' (MacEachren, 1995). This spatially differentiating information is also necessary for the display of quality in a separate map, as this is e.g. the case with reliability diagrams. The two maps can be arranged either next to each other, or can be presented alternating as in the case of an electronic display. If visualising quality information for several datasets in comparison, one should only make use of geodata files of the same data type.

The characteristic of the project described in this paper is the processing and handling of both numerous and varied geodata types, e.g. scanned topographic maps, satellite imagery, vector layers or GPS readings. For this reason visualising geodata quality differences within the mapped extent per dataset would be far too ambitious. This is especially so considering the collection of data depicting former time stages as the exact circumstances of map creation are often simply not known. Therefore a visualisation in the form of a diagram next to the map is the only feasible option. However, to give a single overall quality statement per geodataset would not only be rather disappointing but would also not suffice for the user. Splitting the illustration of quality into several parameters provides a more detailed overview of the quality of a geodataset.

4. DECIDING ON QUALITY PARAMETERS TO DESCRIBE THE GEODATA USED

In this paper the term data quality does not refer to error as the opposite of accuracy. We do not use the term error, because it can have different meanings (see Zhu, 2005). 'Fitness for use' is not an issue either although we are aware that unsuitable data can lead to wrong analysis results. Uncertainty can be seen as an overall term for data that is not an absolute exact image of the objects in reality. Thus, a very high data quality requires a low uncertainty.

The term data quality can be split into different aspects, all contributing to the quality of a geodataset. One has to be familiar with these aspects before their visualisation can be tackled. In literature five parameters are often mentioned: positional accuracy, attribute accuracy, logical consistency, completeness and lineage (e.g. van der Wel et al., 1994; Slocum, 1999). Comber et al. (2006) name them the 'big 5'.

Lineage describes the development of the dataset to the current state (Slocum, 1999). Although there may be several steps to the actual state, we only pay attention to the state or data source type before its integration in our GIS, because quite often more of the dataset's history is not known to us. Positional accuracy deals with the difference of the geodata object to its true geographic position. This can also include the third dimension, such as the accurate height of a mountain (Slocum, 1999). By

geodata quality parameter	what is judged?	additional information
- temporal information (TI)	difference between year on map/geo-dataset and year of content, plus knowledge of these dates	year on map/geodataset
- lineage (Li)	reliability of original (parent) dataset	data type of original (parent) dataset
- positional accuracy (PA)	the georeference of the objects	scale or resolution
- attribute accuracy (AA)	quality of attributes	number of attribute classes
- completeness (Co)	not a judgement, but percentage of completeness	referring to forest boundaries or to study area
- logical consistency (LC)	mainly geometric contradictions in dataset	(none)

Table 2. The six selected geodata quality parameters with their additional information.

attribute accuracy one can express whether the thematic variables were classified in a correct way (Buttenfield & Beard, 1994). A logical, consistent dataset must not have geometric, topologic or thematic contradictions (Navratil, 2006). That includes the relation of the objects in the map to each other having to be correct. In a complete dataset no object must be missing (Slocum, 1999). Due to aspects of generalisation one has to be aware of minimum sizes for mapped objects in order to judge completeness correctly.

Besides these five we consider the temporal information aspect to be of importance too. For a correct joint analysis and interpretation of the geodatasets it is important to know whether or not e.g. the date mentioned on the map corresponds with the content of the map. If the analysis is based on the wrongly perceived date, this can lead to incorrect conclusions on the forest cover. This temporal aspect must not be confused with the often mentioned quality parameter 'currency' which refers to how up-to-date a dataset is (see Navratil, 2006). Instead, our study is to be based on a broad range of geodatasets covering more than the last 100 years in order to investigate the change in forest extent and state due to forest use practices.

The six selected geodata quality parameters are listed and explained in Table 2. Besides a judgement on each of the parameters we put additional information beside each one, e.g. scale or resolution in case of positional accuracy (see Table 2). An exception is the parameter 'logical consistency' that has to make do solely with a judgement. 'Completeness' can be better described than simply with a judgement, i.e. it can be specified as a percentage. Here, as additional information the choice between making reference to the forest boundaries or to the complete study area is given, because for quite a number of datasets information mapped within the forest extents is clearly sufficient. The temporal information has another peculiarity as it consists of three different types of information: the judgement is based on the difference between the year with which the map is labeled and the year of the content. The year on the map is given as additional information. In addition the judgement includes an assessment on how well the actual date is known. In case of a satellite scene, both dates coincide and therefore the reliability judgement is very high, but for old maps it is necessary to address the temporal aspect in a detailed manner.

The parameters can be regarded independently from each other but they are at the same time mutually conditional. When two datasets of different sources are graphically overlayed by means of a GIS, it is certain that imprecision in positional accuracy will appear and will often result from their creation by different organisations for different purposes (see also Longley et al., 2001).

5. JUDGING THE GEODATA QUALITY

Each geodataset was assessed for data quality as a whole against the six different categories as listed in Table 2. For consistency of interpretation all datasets were judged by the same person who had the greatest working knowledge of the datasets and the area.

The lineage scale is related to the purpose of the product used to derive the described dataset for inclusion in the GIS and thus it is an impression of its process of emergence or heritage. The products used can been related to 9 categories (see Table 1 for geodatasets available). The grading is an ordinal scale of 1-5 with 5 being the best. In general, satellite imagery as the source is ranked high, while forestry maps gain a higher grading than forestry sketches. Considering the 132 datasets in total, the 5 gradings show a fairly normal distribution but skewed to its higher end (see Figure 2a).

The positional accuracy was also ranked by 'factors' on a scale of 1-5. Here, the scale or resolution of the graphic enabled a rough ranking as a starting point. For example, datasets of a scale of 1:10,000 (or larger) received a score of 5, while those of 1:1 million or less scored a value of 1. This grading has been further refined by also taking into account knowledge of the georeferencing process or the fitting of features in a visual overlay by means of GIS. In the case of forest logging geodatasets, the ratings for positional accuracy are typically adjusted downwards by a value of 1 in order to compensate for the inaccuracy of the actual logging which is known to often stray beyond the boundaries of marked logging concessions. Typical scales or resolutions of the different geodata types used can be found in Table 1. The statistical graph (Figure 2b) reveals the distribution of the scores. Overall more high scores
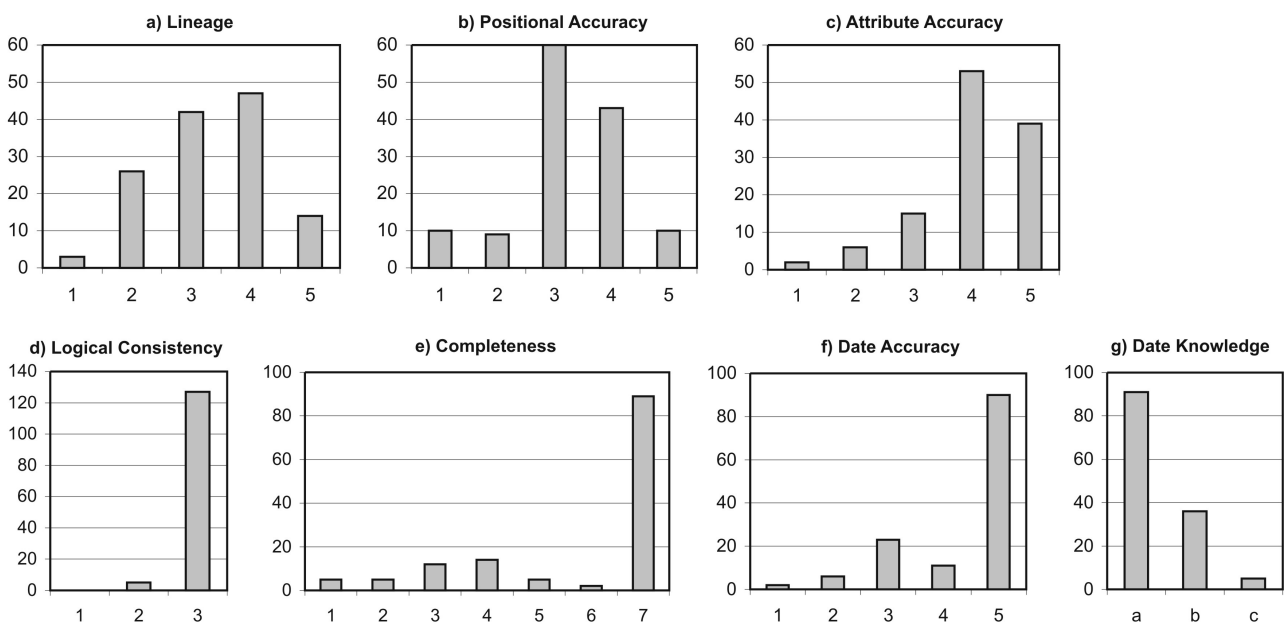
Figure 2. Summary of the geodata quality judgements performed for the 132 datasets on the Kakamega-Nandi area. For a description see chapter 5.
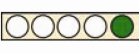
have been given, which reflects the aim of the study, namely to investigate differences in local forest use. A general but worthwhile pattern to mention is that the older the stage represented by the geodataset the lower it has been judged for its positional accuracy.

Attribute accuracy was assessed again on the basis of an ordinal scale of 1 (inaccurate) to 5 (accurate). This judgement is independent of the number of attributes and is solely related to a judgement on the accuracy with which the attributes were assigned. As additional information the number of attributes or datafields are shown and this excludes the default datafields. In the case of imagery, scans of maps or vector datasets showing a single class, for example forest cover only, no attributes or datafields are present. Scanned maps and such vector datasets can nevertheless be judged regarding their attribute accuracy but images not. Figure 2c shows the generally high ranking for our data pool.

Logical consistency normally considers geometric, topological and thematic aspects. In our case every dataset has been carefully checked and corrected as required (in particular for topology), and this quality parameter is predominantely judged on the basis of the correct positioning of the landscape objects in relation to each other. Here a scale of 1 to 3 is used, representing low, medium and high consistency levels. In our case, of all the parameters treated separately, logical consistency is the most difficult to handle by a differentiating judgement. This is because it requires the most detailed knowledge of a dataset which is often not available in the case of geodatasets representing much earlier stages. The judgement has generated a very limited range (Figure 2d) with most datasets scoring the highest class. Only five of the datasets

appear to be inconsistent in terms of positioning of objects in the landscape in relation to each other and four of these represent official boundaries of forests and administrative units. The judgement of completeness is derived from a percentage coverage of either the official forest boundaries (for purely internal forest datasets) or by the percentage represented of the whole 60 x 65 km Kakamega-Nandi study area (for more general datasets). It is the only quality measure that is derived directly from factual numbers although at present in most cases these are only estimated visually. There are 89 datasets that fall into class 7 which represent between 95 and 100% complete coverage of either of the two extents mentioned above (Figure 2e). In those cases in which completeness is based on the forest boundaries most of the geodatasets do not have full coverage, such as a map solely dedicated to South Nandi Forest. While the forestry sketch maps tend to be complete as they relate to isolated forest areas, the datasets resulting from formal forestry maps are the most fragmentary of all since they would often have covered the whole forest in several adjoining map sheets. Their partial coverage here reflects the fragmentary nature of the forestry archives from which most of the forest related datasets were acquired. In the case of the datasets that relate to the extent of the study area the coverage is more frequently complete (52 of 65 as compared to 37 of 67).

Date reliability is measured in relation to the number of years between the date as specified on a map and the date of its actual content. Five classes are created wherein class 5 represents no difference between the two dates, and similarly the awarding of a class 3 score, for instance, reflects a difference of 6 to 20 years. Since some of the maps hold historical information the scale has been set to include those cases of large time spans and

| method		suitable for all parameters	memorable/ overview	adding text possible	under-standable	reference needed	division too exact
	coloured rectangle	ok	ok	good	ok	yes	no
	bar	ok	good	ok (horizontally)	ok	maybe	yes
	slider	ok	ok	bad (e.g. to the right)	good	no	yes
	segment of a circle	bad	good	ok (perhaps a little too tight)	ok	no	no
	traffic light	good, with adaptations	good	ok (e.g. to the right)	good	no	no
	graphic variable	good	ok	ok (perhaps a little too tight)	moderate	yes	no
	line	bad	ok	ok	moderate	maybe	no
	arrow	bad	good, with length	ok	moderate	maybe	no
	plus/minus symbol	ok	good	ok (to the right)	ok	no	no
	circle	bad	ok	ok (perhaps a little too tight)	ok	no	yes

positive assessment

Table 3. Evaluation of general methods used in cartography for visualising geodata quality in diagram form.

thus class 1 represents a discrepancy of at least 100 years (2 datasets). However, for most geodatasets the 'date of the map/geodataset' is consistent with the date of its content (90 of the 132 cases, see Figure 2f). In particular the satellite imagery is, as would be expected, very high scoring. While forestry sketches also score highly since they represent snap-shots in time, the forestry maps are poorly rated here since they attempt to locate multiple data of differing phases of forestry on the same map. A further scale 'a' to 'c' is used to reflect the state of our knowledge of these dates. Thus, 'a' is awarded to a dataset for which the relevant dates are known (being the case 91 times) and 'c' to one for which they are known only vaguely (5 times). The high numbers of datasets for which we have high scores (see Figures 2f and 2g for rankings '5' and 'a') is not surprising since the dates for all the imagery and derived products are well-known. It is the historical anthropological data which has the greatest discrepancy but also vagueness.

6. VISUALISING THE GEODATA QUALITY

As discussed in chapter 3 our geodata's quality is to be visualised in the form of a diagram next to the map. Here six distinct geodata quality parameters have to find space with their specific ways of being judged (see chapter 5). Five of them will be complemented by additional information (see chapter 4). A major aim of the visualisation is that the feature be memorable. But at the same time the details of the information given should be easy and fully to grasp by the user.

6.1 General cartographic methods for visualising geodata quality in a diagram

Table 3 helps to evaluate different general methods widely used in cartography regarding their appropriateness for visualising the geodata quality in diagrammatic form. Here, we dicsuss only the most convenient of these.

A visualisation making use of graphic variables suits all quality parameters. Here, for example, colour saturation could be applied, with a variable number of saturation steps or even a continuous range depending on the differing judgement scales. For a more complex geodata quality judgement, as in the case of temporal information, saturation could be combined with the

graphic variable of colour hue. The implementation via segments of a circle would not fit consistently unless the number of evaluation classes are equal (see Figure 3a). A slider would work well only with more or less continuous judgement scales. Here the introduction of interval markings could enhance its suitability for visualising the distinct parameters. Traffic lights can be easily adjusted to differing numbers of ranking classes. However, the number of coloured lights should preferably be uneven.

The best option for keeping the quality judgement in mind is judged to be the segments of a circle diagram where assessments for all parameters are arranged in a closed shape. Also the options displaying a position along a scale-range are easily memorable (see bar and traffic light). Assistance can also be given by a colour ramp scheme (see slider). Whether the graphics can be accompanied by text depends mainly on the space available, either inside or close to the diagram. In order to demonstrate the connection between the judgement and additional, mostly textual information, the latter should be placed in the immediate vicinity of the diagram. Within a coloured rectangle there is plenty of space to add textual information, even in the case of longer words as used for the lineage information. However, diagrams like the traffic lights require the placement of additional text or figures next to the diagram.

Both the comprehensibility of the diagram and in particular the necessity of showing the minimum and the maximum as reference points contribute to how easily the diagram can be understood. If the ends of a scale are not obvious, as is the case with a coloured rectangle, the user might misinterpret the valuation shown. This is the reason that the slider and traffic lights are so easily understood: here the actual score is presented in relation to the highest and lowest rank. In addition, the traffic lights' interpretation is intuitive, with a red light giving a warning, etc. However, the appearance of the slider encourages the user to presume the rating schemes behind it are continuous. For the same reason the circle should also only be used with caution.

6.2 Combining the quality parameter information in a complex diagram

Having discussed the suitability of general cartographic possibilities we can now discuss the adoption of an effective combined presentation of the six quality parameters in one complex diagram.

The order of the single parameters displayed in the diagram is influenced by a rating of importance for the particular project aim. The quality judgement's level of exactness affects its position in the final diagram too. Temporal information is considered to be the most important parameter because geodatasets including data of the past 100 years and more are used for analysing forest cover change. Considering the long-term nature of the forest research the older documents gain an added status even though their positional accuracy might not be as good. Therefore, temporal information is placed at the top, while the judgement on logical consistency, the least objective and least detailed criterion, is moved to the end.

Four options for combining the different parameters are presented in Figure 3 and are shown simply as graphic concepts, i.e. they are not linked to real geodatasets. In general there are two alternatives for the visualisation. On the one hand a complex representation can be realised with each parameter visualised being customised to its information content. This complex form places emphasis on the most exact communication of information but is less concise (Figure 3d).
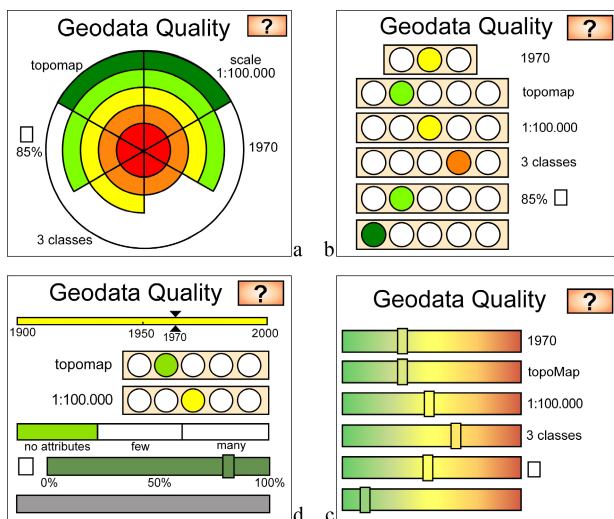

Figure 3. Four alternatives for visualising the six geodata quality parameters in a complex diagram.

On the other hand, the data quality information can be presented in a simpler way where the parameters are visualised similarly (see Figure 3, a-c). These diagrams have the advantage of providing a faster-to-grasp overview that the user can easily keep in mind. From the cartographic point of view it would be best to treat every parameter differently finding the optimal representation for its specific characteristics. But this would require the user to regularly consult a detailed description and would necessitate a lengthy learning period. As the later users will not necessarily be cartographers the simplest visualisation strategy is chosen here.

A major difference between using segments of a circle and the presentation by traffic lights, aside from the arrangement, is the space available to add textual information (see Figure 3, a & b). The segments of a circle make it difficult to position text of differing length. Traffic lights can be presented in a very small size allowing for plenty of space for even longer text lines. Here the text should be positioned next to the diagram and not above or below the traffic light sign, because this would disturb the overview of the constellation of the coloured lights. When making use of sliders (Figure 3c) one has to be aware that the colour representing the judgement has to be visually highlighted in contrast to the colours of the complete colour ramp. When using traffic lights this requirement is easily achieved by simply varying the colour of the particular light.

6.3 The final diagram adjusted to the quality judgement at hand

It has become clear that keeping the visualisation simple is advantageous. Visualisation via traffic lights is seen to be adequate for five of the parameters as their judgement considers five or three ranks. As illustrated by the grey shading in Table 3 this method gained the best overall assessment amongst all the alternatives demonstrated. It is only for the completeness parameter, which does not provide a judgement but a factual measure, that the decision was made to adopt a slider. In order to link the temporal information parameter with three kinds of information the variable crispness as introduced by MacEachren (1995) is planned in order to reflect the state of

Figure 4. The agreed concept for visualising the six selected geodata quality parameters. For explanation of abbreviations see Table 2.

our knowledge on the dates on which the judgement is based. While the colour and position of the light indicates the concept for visualising geodata quality for five distinct judgements on date reliability, three variations in crispness applied to the light reveal the degree of certainty. The redundant expression of data quality information by colour and position prevents interpretation problems due to possible colour-deficiencies.

This final concept for visualising geodata quality for six distinct parameters is shown in Figure 4. The additional, mostly textual information is always placed next to the diagram on the right hand side while an abbreviation for the parameter's name is placed on its left. This arrangement contributes to a consistent and clear overall picture. Using just two letters occupies the minimum space while enabling even a new user to quickly link the correct diagram with each of the six geodata quality parameters. For assistance a help button placed in the upper right corner will lead the user to a comprehensive description of the visualisation concept.

7. OUTLOOK AND CONCLUSION

The concept of visualising quality for varied geodatasets as introduced, described and discussed in this paper is currently implemented in a visualisation tool for displaying and working with spatio-temporal data of the Kakamega-Nandi forest area in west Kenya. The tool will consist of two tabs for changing between a scientific report on forest use history in this area (see Mitchell, 2004) and a display of the geodata available. Hyperlinks in the text will open the map view tab loading relevant geodatasets or centering the map field to a specific location. Further navigation within the map view is enabled by buttons arranged in a toolbar. Several vector datasets can be displayed at the same time and raster datasets can be viewed one at a time. Here a table-of-contents list will provide the required versatility to toggle between the datasets. The display of the geodata quality diagram per dataset is also controlled from here and can be viewed in succession. The programming (see Huth et al., in prep.) is based on XHTML for the text tab and SVG for the map view tab. The database behind is MySQL with access enabled by PHP. Interaction is realised by JavaScript. The tool will not only offer the opportunity to the scientist already familiar with the geodata to gain new insights, but can be of use to a wider audience for simple documentation, presentation of results and provides them the opportunity of working with the gathered data and information. Including a presentation of the quality aspect helps to enhance the understanding of the characteristics and usefulness of the geodatasets and thus allows a judgement of the descriptive text in relation to the geospatial information.

To conclude, geodata quality has been visualised before. Our research, however, provides the opportunity to visualise data quality for a substantial collection of geodatasets of very different origin, data type and quality. The geodata quality judgement carried out in combination with the actual geodata gives a thorough example of real use visualisation in which the geodata quality is not spatially visualised but instead six distinct quality parameters are differentiated. The concept gives special consideration to the temporal aspect of the geodata which covers a period of 110 years. As such it is particularly useful for describing and visualising the quality of geodata collections that include a historical dimension and is readily transferable to such data pools.
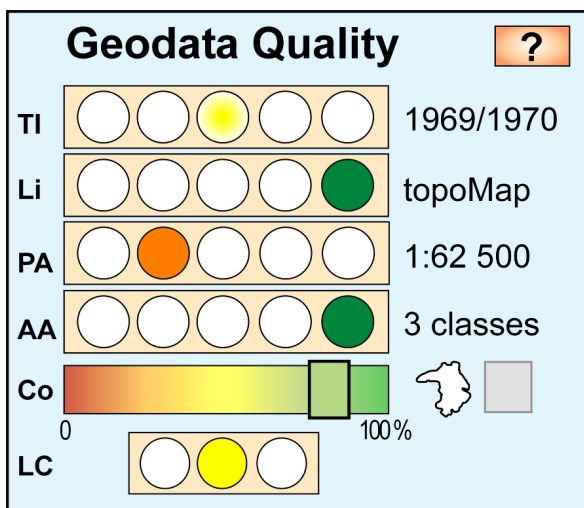
REFERENCES

Buttenfield, B. & M.K. Beard 1994: Graphical and geographical components of data quality. In: Hearnshaw, H.M., D.J. Unwin (Eds.), *Visualization in Geographical Information Systems*, John Wiley & Sons, Chichester, pp. 150-157.

Chapin III, F.S., E.S. Zavaleta, V.T. Eviner, R.L. Naylor, P.M. Vitousek, H.L. Reynolds, D.U. Hooper, S. Lavorel, O.E. Sala, S.E. Hobbie, M.C. Mack & S. Diaz 2000. Consequences of changing biodiversity. *Nature*, 405, pp. 234-242.

Comber, A.J., P.F. Fisher, F. Harvey, M. Gahegan & R. Wadsworth 2006. Using metadata to link uncertainty and data quality assessment. In: Riedl, A., W. Kainz, G. Elmes (Eds.), *Progress in Spatial Data Handling, 12th International Symposium on Spatial Data Handling*, Springer Verlag, Berlin-Heidelberg, pp. 279-292.

Devillers, R., Y. Bédard & R. Jeansoulin 2005. Multidimensional management of geospatial data quality information for its dynamic use within geographical information systems. *Photogrammetric Engineering and Remote Sensing*, 71(2), pp. 205-215.

Fisher, P. 1994. Animation and sound for the visualization of uncertain spatial information. In: Hearnshaw, H.M., D.J. Unwin (Eds.), *Visualization in Geographical Information Systems*, John Wiley & Sons, Chichester, pp. 181-185.

Huth, K. 2007. *Entwicklung eines prototypischen, SVG-basierten Tools zur Visualisierung von Geodaten für das Waldgebiet Kakamega-Nandi in Westkenia unter besonderer Berücksichtigung ihrer Qualität*. Unpublished Diploma thesis, Studiengang Kartographie und Geomatik, Hochschule Karlsruhe – Technik und Wirtschaft.

Huth, K., O. Schnabel & G. Schaab in prep. SVG-based visualisation of geodata quality. Taking the Kakamega-Nandi forest area as an example. In: *Proceedings of the XXIII International Cartographic Conference 2007*, 4-10 August 2007, Moskow, Russia.

Köhler, J. (2004) Was hat Biodiversitätsforschung mit ‚nach-haltiger Nutzung' zu tun? *Tier und Museum*, 8(3), pp. 82-91.

Longley, P.A., M.F. Goodchild, D.J. Maguire & D.W. Rhind 2001. *Geographic Information Systems and Science*. John Wiley & Sons, Chichester.

Lung, T. 2004. Landbedeckungsänderungen im Gebiet "Kakamega Forest und assoziierte Waldgebiete" (Westkenia) – Multispektrale Klassifikation von Landsat-Satellitenbilddaten und Auswertung mittels Methoden im Raster-GIS. *Karlsruher Geowissenschaftliche Schriften*, A 15, ed. by G. Schaab.

Lung, T. & G. Schaab 2004. Change-detection in western Kenya: The documentation of fragmentation and disturbance for Kakamega Forest and associated forest areas by means of remotely-sensed imagery. In: *ISPRS Archives Vol. XXXV Part B* (DVD), Proceedings of the ISPRS XXth Congress, 12-23 July 2004, Istanbul, Turkey.

MacEachren, A. M. 1995. *How Maps Work. Representation, Visualization, and Design*. The Guilford Press, New York.

MacEachren, A.M., D. Howard, D. Askov, T. Taormino & M. von Wyss 1996. Reliability Visualization System (RVIS). http://www.geovista.psu.edu/publications/RVIS (accessed 10 Dec. 2006).

Mitchell, N. 2004. The exploitation and disturbance history of Kakamega Forest, Western Kenya. *Bielefelder Ökologische Beiträge*, 20, BIOTA Report 1, ed. by B. Bleher & H. Dalitz.

Mitchell, N., T. Lung & G. Schaab 2006. Tracing significant losses and limited gains in forest cover for the Kakamega-Nandi complex in western Kenya across 90 years by use of satellite imagery, aerial photography and maps. In: *Proceedings of ISPRS (TC7) Mid-Term Symposium "Remote Sensing: From Pixels to Processes"*, 8-11 May 2006, ITC Enschede, The Netherlands, ed. by N. Kerle & A.K. Skidmore.

Mitchell, N. & G. Schaab subm. Developing a disturbance index for five East African forests using GIS to analyse historical forest use as an important driver of current land use/cover. In: *African Journal of Ecology*.

Mitchell, N. & G. Schaab 2006. Assessing long-term forest cover change in East Africa by means of a geographic information system. In: *Hochschule Karlsruhe –Technik und Wirtschaft, Forschung aktuell 2006*, pp. 48-52.

Navratil, G. 2006. Data quality for spatial planning – An ontological view. In: Schrenk, M. (Ed.), *Sustainable Solutions for the Information Society, Proceedings of the 11th International Conference on Urban Planning and Spatial Development in the Information Society (CORP)*, 13-16 February 2006, Vienna, Austria, pp. 99-105.

Sala, O.E. & T. Chapin 2000. Scenarios of global biodiversity for year 2100. In: *GCTE News. Newletter of the Global Change and Terrestrial Ecosystems Core Project of IGBP*, 16, pp. 1-3.

Schaab, G., T. Kraus & G. Strunz 2004. GIS and remote sensing activities as an integrating link within the BIOTA-East Africa project. In: *Sustainable Use and Conservation of Biological Diversity – A Challenge for Society. Proceedings of the International Symposium Berlin*, 1-4 December 2003, Berlin, Germany, pp. 161-168.

Schaab, G., T. Lung & N. Mitchell 2005. Land use/cover change analyses based on remotely-sensed imagery and old maps as means to document fragmentation and disturbance for East-African rainforests over the last ca. 100 years. In: *Proceedings of the International Cartographic Conference 2005*, 9-16 July 2005, A Coruña, Spain.

Slocum, T.A. 1999. *Thematic Cartography and Visualization*. Prentice-Hall, Upper Saddle River (New Jersey).

Van der Wel, F.J.M, R.M. Hootsmans, F. Ormeling 1994. Visualization of data quality. In: MacEachren, A.M., D.R.F. Taylor (Eds.), *Visualization in Modern Cartography*, Serie Modern Cartography, Vol. 2, Elsevier Science Ltd., Oxford, pp. 313-331.

Zhu, A.-X. 2005. Research issues on uncertainty in geographic data and GIS-based analysis. In: McMaster, R.B., E.L. Usery (Eds.), *A Research Agenda for Geographic Information Science*, CRC Press, Boca Raton-London, pp. 197-223.