Concept Testing of Some Visualization Methods for Geographic Metadata

Paula Ahonen-Rainio

Helsinki University of Technology, Department of Surveying,
Institute of Cartography and Geoinformatics
PO Box 1200, FIN-02015 HUT
Paula.Ahonen@hut.fi

Abstract. Concepts of metadata visualization have been developed in order to help users in understanding of metadata about geographic datasets. These concepts apply graphic samples and four multivariate visualization methods: scatter plot matrix, parallel coordinates plot, star plots and Chernoff faces. The concepts were tested with users of geographic metadata in the Finnish Defence Forces. The test results indicate that users benefit from graphic samples especially when comparing alternative datasets. When the differences of the datasets are visible users start to question their presumptions about the data. The parallel coordinates plot and star plots support users in comparison of several metadata elements of several datasets simultaneously whereas the scatter plot matrix is difficult method to adopt and does not support a holistic view of datasets. Chernoff faces arouse emotions that may disturb the interpretation of metadata; interpretation of extreme values from the faces is reliable but intermediate values are easily confused.

1 Introduction

While metadata of geographic information are available in increasing amounts their usability becomes a relevant concern. Present metadata services provide users with metadata most often in a textual form though the same users work with geographic data in visual forms. In practise, visual metadata has been limited to sample maps and to maps showing the regions covered by data. Textual metadata can express some characteristics of data very exactly but may be very limiting is other aspects. For example, visualization of metadata about several datasets at a time could increase remarkably the usefulness of geographic metadata.

This paper reports on an early stage usability testing of two kind of visualization concepts of geographic metadata: the graphic sample of data and visualization of metadata by four multivariate visualization methods.

The international metadata standard for geographic information (ISO/FDIS 19115:2003) defines one graphic metadata element, the browse graphic, to illustrate a dataset. The standard leaves the content of the illustration open but a natural option is that the browse graphic provides a graphic sample of geographic data as a part of metadata. The expression power of a graphic sample depends on the nature of the geographic data that the sample illustrates. The main distinction is whether the geographic data are intended for visual use or computational use. It is unknown how much and what kind of meta-information different users can derive from graphic samples.

Multivariate visualization methods make multiple variables of numerous objects visible at a time allowing exploratory studies of the data. Several multivariable methods are presented for information visualization (Card et al. 1999, Spence 2001); among them are methods that allow identification of individual objects from the displays and can be applied to a relatively small number of objects, which is a typical case with metadata. Ahonen-Rainio and Kraak (2003) propose a visualization environment with four multivariate visualization methods for geographic metadata. Visualization environments make use of multiple linked views in order to support users in obtaining an insight to the information. Therefore, different methods may be employed in parallel but a high degree of interaction is required.

When the goal of visualization is to increase the usability of geographic metadata users' acceptance of the visualization methods is crucial. In the context of geovisualization, the issues of usability and cognition have been recognised as factors that effect on how much benefit may users derive from the visualization techniques or whether the techniques are utilized at all (Slocum et al. 2001). Various usability studies of geovisualization techniques have been reported recently (e.g. Andrienko et al. 2002; Furhman and MacEachren 2001).

Users of metadata typically study metadata only occasionally and, therefore, they cannot be assumed to be motivated to put effort on learning to interpret complex visualization methods. Researchers and other users who are strongly involved in clarifying the characteristics and nature of data are the main uses of information visualization methods. Though users of metadata share the same need for insight, they consider metadata rather a supportive mean than the main interest of their work. Therefore, we must ensure that the information visualization methods applied are easily adoptable to occasional users of metadata and do not require too much involvement (as methods) from their users.

In this study, users are involved in an early design stage of metadata visualization. By communicating the ideas of metadata visualization to users we can test the concepts and gain understanding about:

- how would users benefit from visualization of metadata and which kind of visualization methods they would consider most useful in various stages of the process of using metadata
- which are the significant design factors of visualization of metadata in respect to utility and usability (Nielsen 1993)
- how intuitive and reliable users consider their interpretation of metadata presented by different visualization methods

This study is a part of research on visualization of geographic metadata to help users in understanding the information expressed by metadata. The results of the concept testing can be taken into account when developing the visualization environment and designing the future (quantitative) studies.

2 Testing of the Visualization Concepts

This study is composed of two parts according to the two visualization concepts that were tested:

- graphic samples of vector data and
- four multivariate visualization methods: scatter plot matrix, parallel coordinates plot, star plots and Chernoff faces.

In the first part, the concept is very simple vector graphics. Though such line drawings are familiar to any user of geographic information their utility in the context of metadata has not been studied. The interest was to find out what can users detect from graphic samples of vector data and what attracts their attention, whether they consider graphic samples useful as a part of metadata, and whether a reference image can help users in interpretation of characteristics of sample data.

In the second part, the concept is based on multivariate visualization methods that are typically used for visualizing huge numbers of data items in an exploratory context. The methods as such are not new but their usefulness to metadata has not been tested. The aim of testing was to get an idea how useful the users of metadata find these methods. Furthermore, the idea of exploratory comparison of datasets was introduced to users with the aim of reaching preliminary understanding about how the users would use the various methods during the process of studying metadata.

Concept testing is a communication process between designers and potential users about the form, function and features of a product (Ulrich and Eppinger 2000). There are two aspects: communicating the ideas to the test users and collecting feedback from them. Several methods and techniques can be used for communicating the ideas, like use scenarios, storyboards, paper prototypes and physical mock-ups (Kuutti 1995; Erickson 1995). Erickson (1995) reminds that roughness of a prototype attracts more comments on the idea than a final-looking prototype that draw the attention of users from the principle idea to the details of the design.

In this study the concepts were communicated to users by static displays. Interactivity is a core element of an exploratory visualization environment but in this study the interactivity was not implemented but introduced to test users by use scenarios. Users were told about the principle techniques of interaction and how they could be used with the multivariate techniques. The aim was to ensure the feasibility of the metadata visualization methods with users before implementing them in an interactive prototype.

Collecting of feedback from users can be done, for example, by an interview, a focus group discussion or by recording the thinking aloud of users when they study the ideas or use a prototype. It is important to hear the voice of users; the subjective and ambiguous views of the users may reveal valuable consideration in the design process (Erickson 1995) rather than the generalized, “objective” results of quantitative research methods (which may have their role in later stages). In this testing the users were thinking aloud when studying the displays. In addition, a semi-structured interview was carried out throughout the testing.

For the validity of usability testing, it is important that the test users represent the potential users. But even then, individual differences between test users cause a problem. The number of test users that give sufficiently reliable results depends on the usability attributes that are tested and the goal of testing. For example, Nielsen

(1993) proposes that already 3-5 users reveal the majority of usability problems for practical development purposes and Beyer and Holtzblatt (1998) propose 10-20 users for collecting information about overall working processes and 6-10 users when studying requirements for detailed user interfaces. This study was carried out with 12 test users from the Finnish Defence Forces. All the test users are working with geographic data, either as application developers, data administrators or application users. Their working experiences with GIS vary from two years to over 15 years.

2.1 Test Setting

User testing of the visualized metadata was carried out in December 2002 - January 2003. Users made the test individually. After an introduction the moderator gave the test materials that were organized in sets of PowerPoint slides on a portable PC and instructed the user how to progress.

Each test user studied the static displays of visual metadata. They progressed at own pace, sometimes returning to previous displays. The moderator encouraged users to thinking aloud if they silently stopped to study a display. There were questions written on displays to encourage the users in commentating. The questions were at a general level in order not to limit too much the thinking of the users; for example, what information users could derive from the display or how users would compare successive displays. In addition, the moderator made further questions on the basis of comments of the users. The more the users thought aloud the less active was the role of the moderator. Each test session took about one hour and was audio recorded. Afterwards the moderator transcribed the recordings for analysis.

2.2 Testing of the Graphic Samples

The graphic samples were produced from three vector datasets: two road datasets and a datasets of coastlines. The road datasets originate from different data suppliers and have slightly different definition of the content. The coastline data originates from a 1:250 000 map.

The sample data is from the town of Kotka on the coast in the South Eastern Finland. The sample area of about 2 km x 2km includes a dense housing area with streets, an uneven seashore and a small forested outdoors area. The area was selected randomly only concentrating on the versatility of feature types. It appeared that many curiosities worth of detecting were though included in the sample data. The graphic samples presented the pure data and had no added marks that could have signalled the curiosities or potential errors of the data.

In addition to graphic samples, reference images were included in the displays. The role of the reference images was to represent the real world, and the idea was to support users in judging characteristics of the data by allowing comparison with the reference. Because the real world is continuous and infinite in its details the representation can never be more than an approximation. Therefore, a reference image at its best can only give an approximation of higher grade than what the data does. The reference images used in the test were a city map, a topographic map and an orthophoto (see Fig. 1 and 2). They have different limitations in respect to the level of details and currency.

The graphic samples were displayed individually, in parallel to each reference image and overlaid on the orthophoto. In addition, the two road data samples were displayed in parallel to and overlaid on each other, and the coastline data sample was displayed with a sample of coastline data from the national topographic datasets that corresponds to map scale 1:20 000 – 1:50 000. Examples of the test displays are given in Figures 1 and 2. The original displays were in colours.

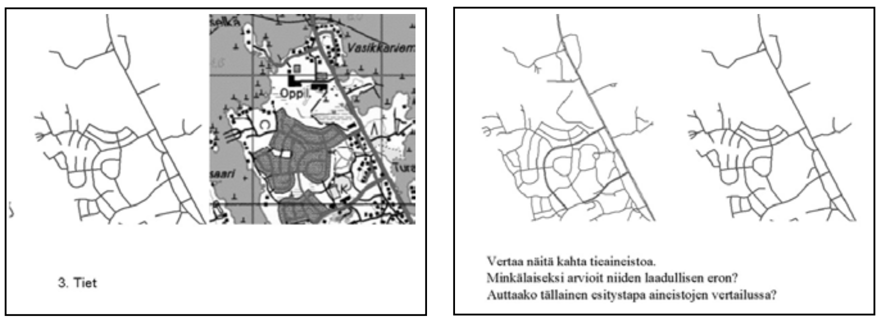

Fig. 1. Examples of the graphic sample displays. On the left, the sample of the first road dataset is in parallel to the topographic map. On the right, the samples of the two road datasets are in parallel. The texts in the displays are in Finnish.

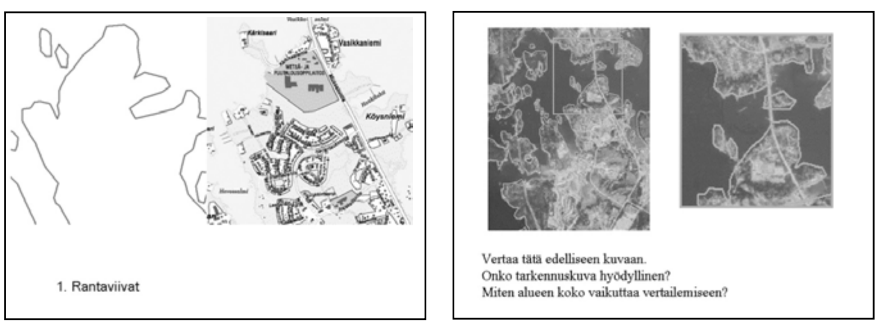

Fig. 2. On the left, the sample of the coastline dataset is in parallel to the city map. On the right, the sample of the coastline is overlaid on the orthophoto and presented in two different scales.

Users were asked in general to study the displays in order to construct knowledge about the data of the graphic samples. They were asked to consider what meta-information they could derive from the individual graphic sample and what more they could derive when a reference image was provided. Also they were asked to compare different reference images in respect to utility.

2.3 Testing of the Multivariate Visualizations

For the testing of the four multivariate visualization methods, a set of metadata was compiled. The set was composed of five metadata elements for eight road datasets that are available for Helsinki region. The metadata elements selected for the test were updating frequency, scale (or reference scale), geometric structure, price and number of geometric objects (i.e. road segments).

The domains of corresponding (ISO/FDIS 19115:2003) metadata elements were applied, and the metadata values were manipulated as ordinal data for the displays. Some values were missing, which reflects the metadata being incompletely available. The ordering of the metadata values would actually be a task of the user but in the testing the order was predefined and documented as a part of the use scenario.

Four multivariate visualization methods were selected for testing on the basis of their popularity and their applicability to the case of metadata (Ahonen-Rainio and Kraak 2003). The methods are explained below and examples of the test displays are given in Figures 3-6. The colour coding of the datasets is given in each display and the variables are explained (in Finnish). The quality of the displays is heavily reduced by the grey scale printing.

Scatter plot matrix. A scatter plot matrix is an extension of the conventional approach to the visualization of bivariate data where a two-dimensional plot presents one variable against the other. In a scatter plot matrix all pairs of axes are presented as scatter plots and these are then arranged in a matrix. The scatter plot matrix display is shown in Figure 3.

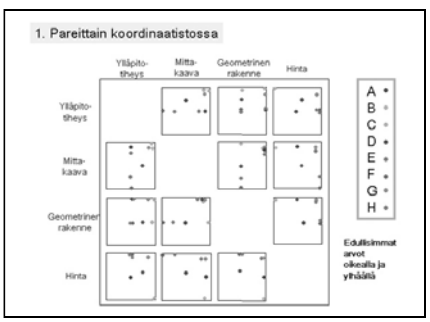

Fig. 3. Four elements of the test metadata are displayed in a scatter plot matrix.

Parallel coordinates plot. In a parallel coordinates plot (Inselberg 1985) the variables are represented as coordinate axes arranged in parallel to each other. Values on a coordinate axis are typically scaled so that the ends of the axis represent the maximum and minimum values of the related variable. A parallel coordinate plot representing metadata plots each dataset as a line going through the variable values on the corresponding axes. The method treats the whole set of variables equally because each of the variables has the same graphic representation. However, the ordering of the axes influences the nature of the plot and so possibly its interpretation. Therefore, it is important that users can control the ordering during interactive exploration so that the variables of interest can be studied in adjacent axes.

Parallel coordinates plots were produced of the five metadata elements. One of the displays showed metadata of all the eight datasets; in another display the two "best" datasets were removed (Fig. 4).

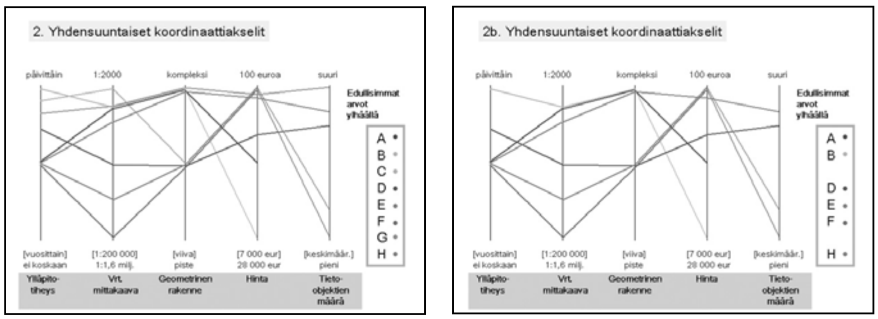

Fig. 4. The metadata are displayed in a parallel coordinates plot for all the eight datasets (left) and for six datasets excluding the two "best" datasets (right).

Star plots. In a star plot (Chambers et al. 1983), the axes for variables radiate in equal angles from a common origin. When visualizing metadata, there is a star plot for each dataset. A line segment is drawn along each axis starting from the origin and the length of the line representing the value of the related variable. An outline connects the end points of the line segments, thus creating a star shape. The ordering of the individual axes influences the shape of the star plots, as does the ordering of the axes of parallel coordinates plots. A test display of star plots showed the eight datasets; in another display the stars were plotted with axes in full length (Fig. 5). A third display showed four star plots overlaid for comparison purposes.

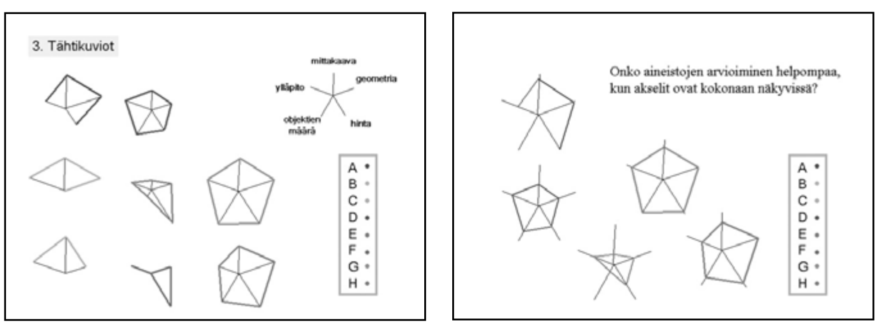

Fig. 5. The eight star plots on the left show the datasets with all the five metadata elements. On the right, the star plots are with axes in full length.

Chernoff faces. Chernoff faces (Chernoff 1973) are multi-part glyphs in the shape of a human face and the individual parts, such as eyes, ears, mouth and nose represent the data variables by their shape, size and orientation. The idea of the use of the faces is related to the fact that humans easily recognize faces and can see small changes without problems. Chernoff faces handle each variable differently. Because the features of the faces are different in importance the way in which variables are mapped to the features is critical. Therefore, the user should be allowed to set the priority for the metadata elements for the visualization. Spence (2001) reminds that icons have an advantage over textual representation when there is a semantic relation between the icons and the task. In order to gain this advantage, the icons in the test were to show a relation between the favourable metadata values and satisfied face features. A test display of Chernoff faces (Fig. 6) shows only three of the metadata elements.

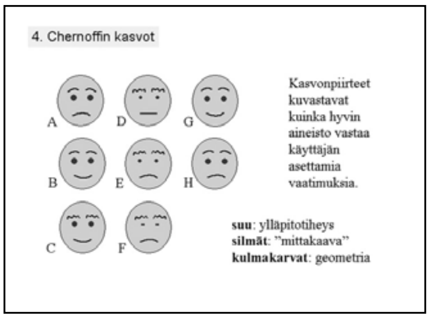

Fig. 6. Three metadata elements are displayed in Chernoff faces.

Users were first introduced to a use scenario telling that they were to select the dataset(s) that would best meet their needs. The scenario fixed the criteria for the best dataset; the metadata values were ordered for the visualization in accordance with the scenario. The users were told that in the real situation the working environment would be interactive and they would themselves first select the metadata elements, their priority and the preferred order of the values of each metadata element. The users were then introduced to each of the methods at a time and asked to select from the display the most favourable dataset(s). Most importantly, they were asked to think aloud while they studied the displays.

The original displays are in colours. In the test colour was an easy way for users in their speech to refer to individual datasets. In an interactive environment the identification can be done differently and, consequently, some problems relating to selection of colours can be avoided.

During the test the moderator discussed with the test users about typical techniques implemented in exploratory visualization environments, like ordering of axes of a parallel coordinates plot, multiple linked views, brushing and filtering that have been demonstrated also in geovisualization environments (e.g. Haslett et al. 1990; MacDougal 1992; Dykes 1995). In this way users got a rough idea of interactivity and they commented how they would use the possibilities in their work with metadata. At the end of the test session users were asked about their opinion of visualization of metadata in general and the situations where they could benefit of visual metadata.

3 Results

The main results of the concept testing are reported here. Many details as "the voice of users" are documented separately for consideration in the further design studies.

3.1 Graphic Samples of Vector Data

The test showed that users can derive very little meta-information by looking at an individual graphic sample of vector data. But when samples of two datasets representing the same theme are shown together users awake to study the differences. The test clearly indicates that users rely on their presumptions about geographic data; they assume good quality if the data displayed looks reasonably clean and do not question the details of the semantics, spatial accuracy, completeness or consistency of the data. When users see differences between datasets they become aware of the possibility that the data are not what they assume them to be.

Comparison of Two Samples. The test users first studied displays of a road dataset that looked fine but did not include all the smaller roads. Only few users commented that some minor roads were missing when they compared the data with reference images. When the second road dataset was shown in parallel to the first one the difference in the level of details was clearly visible and users started to comment the lack of minor roads in the first dataset. When the two graphic samples were overlaid the differences could be seen in detail. The second dataset included many more roads, but some roads that were included in the first dataset were missing from the second one. Half of the test users considered this as a confusing piece of meta-information whereas the other half just disregarded that observation.

Comparison of a sample to a reference image. Comparison of graphic samples to a reference image helps users in deriving meta-information from the data display but the selection of an optimal reference is a complex issue. Those test users who are working a lot with aerial photographs preferred the orthophoto as a reference image. But the orthophoto was useless for those test users who are not accustomed to photo interpretation. They preferred a map. The reference map should present the theme in question most clearly. All the users preferred the topographic map as a reference of the road data; the roads were difficult to detect in the city map because the road names made the image fuzzy and no distinguishable colour was used. Several test users mentioned that the topographic map is a natural reference for military people as they use it frequently and consider it as a kind of standard. However, the city map appeared to be the better reference when the shoreline was studied. The shoreline in the city map was a continuous line in bright colour without disturbing elements whereas, in the topographic map, rock symbols cut in the shoreline here and there and occasionally the shoreline changed from a continuous to a dotted line indicating reeds on the shore. This concludes that the theme of the data and the expertise of the users determine what is the most suitable reference in each case. The colours (or any other visual variables) used in representing the theme in the graphic sample and in the reference image shall be comparable or, preferably, the same.

Users' presumptions about Geographic Data. Thinking aloud revealed that users have many presumptions about geographic data. Part of these could be called as naïve meta-information. Many users gave comments like "the road data with more details is more accurate" though there was no explicit information about the accuracy. Some users also thought that road segments in the data that could not be found in the reference image indicated that the data is more current than the reference image. Nobody expressed a concern that the data might include roads that do not exist in the reality. Some users, however, considered that the second road dataset with more detailed data must be "raw data" still requiring consistency checking and geometric cleaning because the lines were "uneven". (The sample data was from a final dataset.) Many users commented that difference in their age must be the reason for the differences between the data and a reference image.

The most experienced users of geographic data made more consistent observations than the least experienced users. But the consistency of the comments did not completely follow the experience (in years) of the test users. More importantly, the attitudes of the users towards geographic data seem to matter and personal differences seem to have an impact on how concentrated observations users do. The test situation of course differs from real life when the user of metadata may have a stronger motivation and aspects for urgent consideration in mind.

3.2 Multivariate Visualizations

The overall attitude of the test users towards the multivariate visualisation methods was positive. Still there were differences in the user comments about the methods. The parallel coordinates plot and star plots gained the most positive reactions. Several users found the scatter plot matrix a difficult method to adopt. In addition, it does not support a holistic view of datasets. Chernoff faces aroused emotions that may disturb the interpretation of metadata.

Scatter Plot Matrix. The scatter plot matrix appeared to users as a difficult method. Only one fourth of the users gave the method their unreserved approval. Half of the users considered the method very complex and non-easily approachable, even though they could detect the best datasets from it in a reasonable time. Two users lost their motivation after some attempts of interpreting the display. The strategies of studying the display varied among the test users. Only few users seemed to gain benefit of the expression power of the whole matrix. This emphasise the importance of guidance in exploratory strategies.

The scatter plot matrix is not a favourable method for occasional users nor does it support building of a holistic view of a dataset. Though each scatter plot is simple to read as such, users have to keep in mind the various plots and are not really supported in constructing knowledge of the whole dataset. Pair-wise relationships that the scatter plot matrix presents explicitly are not the main interest in comparison of metadata. Redundant information in the matrix makes the display quite crowded.

Users easily ignored datasets that were missing in a scatter plot because of a missing value. It is important to consider how interaction would help in these cases.

Parallel Coordinates Plot. The parallel coordinates plot provides a holistic view of datasets as several metadata elements can be shown in the same plot. The test users considered a parallel coordinates plot more explicit and faster to interpret than a scatter plot matrix. Half of the test users clearly preferred a parallel coordinate plot to a scatter plot matrix. Test users also made remarks that perceiving the range of values of each metadata element was intuitive in the parallel coordinates plot.

When there is a dataset with all values close to the satisfactory ones (as in Fig. 4 on the left) it is very easy to recognise the line of that dataset in the plot. When no unambiguously good dataset is available ones (as in Fig. 4 on the right) the interpretation of the plot is not as straightforward, but the test users felt confident when selecting the best datasets. Users commented on the complex outlook of the plot when there are several crossing lines and favoured the idea of reorganizing the axes. One third of the test users comment that in some stage they would remove those datasets from the plot that would not in any case satisfy their needs. Consequently, the plot would become simpler and more easily interpretable.

It is easy to detect in the parallel coordinates plot that values are missing. The axes with values missing where organized in the test plots at the left so that some of the lines ended before the last axis but the lines did not break in the middle. This is not a sustainable solution but other techniques should be considered.

Star Plots. The test users favoured star plots commenting them as an intuitive and expressive method that gives an impression of the “overall goodness” of each dataset. Only one test user considered star plots as an inferior method. Even if star plots give a clear view of each dataset they are not easy to compare if the differences between the plots are minor or in different axes. In addition, it is impossible to attach the names of the axes to the plots, as they would attract the attention from the star shape. Therefore, a study of individual metadata elements would require some effort of tracking and keeping in mind how the axes relate to the metadata elements.

In general, users considered that star plots would best provide the first impression of datasets. Then users would select those datasets they would like to study further and view them in a parallel coordinates plot. An overlay of star plots was also shown to users but they considered that for a more detailed study it would not really add anything to the parallel coordinates plot. Rather it would easily result in overplotting.

Two users commented that the star plots showed the missing values more clearly than the other methods. However, in the comparison most of the users completely ignored those star plots that had values missing. A display with axes in their full length balanced the star plots so that it was easier to distinguish which values were missing and which values were dissatisfactory (i.e. close to the origin).

Chernoff Faces. The Chernoff faces aroused emotions among the test users. Some of them were very amused when studying the faces; some found the faces irritating and not a suitable method for providing serious information. Regardless of their emotional reactions, users could easily identify those faces that visualized a coherent set of either satisfactory dissatisfactory metadata values. But when the faces represented a mixture of satisfactory and dissatisfactory values or values between these extremes the users made confusing interpretations. This happened even though there were only three variables.

In addition, it appeared that the features interact. The even eyebrow curve that indicated a satisfactory value and was easily attached to that value in a happy face was interpreted as a dissatisfactory value when in the face where the mouth was a downward curve (see Fig. 6 faces A and B). In general, the eyebrow caused most confusion whereas the mouth curvature was interpreted without mistakes.

Users seem to interpret the faces very quickly. A face is an icon that gives an immediate impression that the user does not question further. Therefore, the Chernoff faces should be applied only when sure of the correct interpretation.

4 Discussion

The test users outlined the process of studying metadata quite unanimously: after searching the potential datasets by theme users would first select the metadata elements for multivariate visualization; then the metadata would be visualized, first by star plots and then by a parallel coordinates plot – perhaps users would filter out some of the datasets at some point; by exploration, users would select one or two datasets that would then be visualized by graphic samples for the final evaluation. The users considered that graphic samples are useful for users at any skills level. They considered multivariate visualizations very useful if alternative datasets are available; most test users claimed that they never be in such a situation, mostly because of the coordinated use of geographic data in the Defence Forces. Though the process seems to be very straightforward, there are many issues that should be studied before visualization environment could be developed on a solid basis.

4.1 About the Graphic Samples

Attention of users. The test pointed out that user can derive very little information from an individual graphic sample and users' skills vary in detecting important details. Somehow users' attention should be draw to important details. A checklist of items that might be relevant for each type of data could give support for users. A checklist would be transparent and, as such, support novice users in learning of geographic information. Production of graphic samples could be automated, as no signals would be added to the actual graphics.

Selection of samples. For the test the sample area was selected randomly after a rough decision about the location. The resulting sample however included a rich variety of details and proved to be a valuable example in the testing. It is impossible to draw a conclusion that a randomly selected sample would always be satisfactory but it should be studied what would be the sufficient parameters for automated sample selection. These parameters would determine the location, extent and presentation scale of samples. If the phenomenon varies across the entire dataset there should be a set of representative samples that cover the variation of the data.

Visual design of samples. Visual design of the samples is of outmost importance. The interpretability of and the impression that users get from a sample depend on the selection of visual variables. For example, in a pilot test before the actual testing a graphic sample of the coastline without a reference image was removed. The display was unintelligible to the pilot test user because there was only the coastline without a hint about what is water and what is land. It had been possible to colour the water area in blue if the data had comprised water and land areas. But if the data provided the coastline only the coloured water would give users a completely wrong impression of the structure of the data. Skilled users might benefit from a possibility to modify the variables interactively. The design principles of graphic samples are an issue of further studies.

4.2 About the Multivariate Visualizations

The test confirmed that users of metadata prefer easily adoptable visualization methods. The preferences of the users were clear among the tested methods. Other information visualization method should not still be excluded as long as they meet the criteria of metadata visualization. At the concept level the test users seemed to adopt easily the ideas of interaction that is essential in the exploratory visualization. The user interface for interaction must be as intuitive as possible.

It is important that visualized metadata are provided to users together with textual metadata. Visualization is an efficient way of providing an overview and information about relationships, trends and outliers, but other means are needed as well. For example, textual information about the spatial accuracy and currency of geographic data is required in parallel to graphic samples.

4.3 About the Concept Testing

When test users think aloud consistently that reveals a good insight to their attitudes and experiences and to the details that bother them or that they ignore. This kind of qualitative testing gives a wider perspective to a designer or a researcher than quantitative studies alone could give. For example, it was possible to measure from the audio recordings how long it took from the test users to find the "best" datasets in the scatter plot matrix. The performance time was not really long and there was very little variation between the users. However, some users find the method quite good whereas some other users considered it very frustrating. A good result in performance time did not correlate with the experience of confidence or convenience.

Listening to test users can be very stimulating. On the other hand, evaluation of general validity or reliability of the test results is difficult. However, some conclusions can be drawn. For example, the test revealed differences in working styles: some users are concentrating in exploration whereas some are looking for quick results. The differences in expertise with geographic information explain the differences in user preferences only partly. Therefore, more attention should be put on studying the different users of metadata but in any case the metadata visualization environment shall provide flexible tools. At the same time, users gave quite consistent feedback, for example, on usability of scatter plot matrix and utility of graphic samples.

The number of 12 test users appeared to be sufficient in this test. After the first 3-4 test users the general findings were more or less established but the further test users gave remarks on more specific issues on account of their special knowledge or personal characteristics. For example, a very critical test user and a very busy one gave ideas about how those kinds of users would approach visual metadata. An application user with short experience had difficulties in considering the role of reference images at the level of metadata. A researcher in the navy categorized the coast types and a system developer listed typical errors in geographic data.

5 Conclusions

An early stage user testing was carried out on concepts of metadata visualisation. The concepts were a graphic sample of vector data and multivariate visualization by a scatter plot matrix, a parallel coordinates plot, star plots and Chernoff faces. Test users found the metadata visualization concepts useful though there were differences between the usability and utility of the multivariate visualization methods.

The results of the testing propose that an exploratory visualization process on metadata would start in comparison of several datasets by star plots and a parallel coordinates plot. When the most potential dataset(s) had been identified graphic samples would support the user in finalizing the evaluation of the data. The study confirmed some preliminary ideas of the methods but it also raised several issues that need to be studied further. The expression power of graphic samples should be studied as well as the possibilities of supporting (novice) users in the evaluation process. The selection and design of the graphic samples should be modelled so that production of samples could be automated. Interaction is a vital component of an exploratory visualization environment and it should be implemented with the approved visualization methods in a prototype that can be used in further studies. Research on the user differences in knowledge construction by geographic metadata would lay a basis for studies on metadata visualization.

Acknowledgement

The study was a part of a research project on metadata of geographic datasets financed by the Topographic Service of the Finnish Defence Forces. The financial support as well as the collaboration of the test users is appreciated.

References

Ahonen-Rainio, P., Kraak M. J. Towards multivariable visualization of metadata of geographic information. Submitted: Dykes, J., MacEachren, A. M., Kraak M. J. (eds). *Exploring geovisualization*. Amsterdam, Elsevier, 2003.

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J., Kretchmer, U. Testing the Usability of Interactive Maps in CommonGIS. *Cartography and Geographic Information Science*, 29(4): 325-342, 2002.

Beyer, H., Holtzblatt, K. *Contextual Design. Defining Customer-Centered Systems*, Morgan Kaufmann, San Francisco, 1998.

Card, S. K., MacKinlay, J. D. and Shneiderman, B. *Readings in Information visualization. Using Vision to Think,* Morgan Kaufmann, San Francisco, 1999.

Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. *Graphical Methods for Data Analysis*. Wadsworth Press, Belmont (CA), 1983.

Chernoff, H. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68: 361-367, 1973.

Dykes, J. A. Cartographic Visualization for Spatial Analysis. In *Proc. International Cartographic Conference*, Barcelona. Pages 1365-1370, 1995.

Erickson, T. Notes on design practice: Stories and prototypes as catalysts for communication. In Carroll, J. M. (ed.) *Scenario-based design, envisioning work and technology in system development*. Pages 37-58, John Wiley & Sons, 1995.

Furhman, S., MacEachren, A. Navigation in desktop Geovirtual Environments: Usability Assessment. *Proc. International Cartographic Conference*, Beijing. Pages 2444-2453, 2001.

Haslett, J., Wills, G. J., Unwin, A. R. SPIDER: An interactive statistical tool for the analysis of spatially distributed data. *International Journal of Geographical Information Systems*, 4(3): 285-296, 1990.

Inselberg, A. The Plane with Parallel Coordinates. *The Visual Computer*, 1(2): 69-91, 1985.

ISO/FDIS 19115:2003, Geographic information – Metadata, Final Draft International Standard, *International Organization for Standardization*, 2003.

Kuutti, K. Work process: Scenarios as a preliminary vocabulary. In Carroll,_J. M. (ed.). *Scenario-based design: Envisioning work and technology in system development*. Pages 19-36, New York, John Wiley, 1995.

MacDougal, E. B. Exploratory analysis, dynamic statistical visualization and geographical information systems. *Cartography and Geographical Information Systems*, 19:237-246, 1992.

Nielsen, J. *Usability engineering*. Academic Press, Boston, 1993.

Slocum, T,. Block, C., Jiang, B., Koussoulakou, A., Montello, D., Fuhrman, S., Hedley, N. Cognitive and Usability Issues in Geovisualization. *Cartography and Geographic Information Science*. 28(1): 61-75, 2001.

Spence, R. *Information Visualization*. Addison Wesley/ACM Press Books, Harlow, 2001.

Ulrich, K. T., Eppinger, S. D. *Product Design and Development*. 2nd ed., McGraw-Hill, 2000.