THE CARD SORTING METHOD FOR MAP SYMBOL DESIGN

R. E. Roth [a], B. G. Finch [a], J. I. Blanford [a], A. Klippel [a], A. C. Robinson [a], A. M. MacEachren [a]

[a] GeoVISTA Center, Dept. of Geography, Penn State University, 302 Walker Building, University Park, PA (USA)
(reroth, bgf111, jib18, klippel, arobinson, maceachren)@psu.edu

Commission VI, WG VI/4

KEY WORDS: card sorting, map design, map symbology, emergency mapping, symbol standards, user-centered design, usability engineering, ANSI INCITS 415-2006

ABSTRACT:

Here, we describe the potential utility of the card sorting method for structuring and refining map symbol sets. Card sorting has been proposed as a method for delineating categories by researchers and practitioners in a variety of disciplines due to its ability to identify and explicate real or perceived structures in an information space; however, there is little reported application of card sorting within Cartography. To span this gap, we offer a framework that prescribes the appropriate methodological parameters for card sorting according to the stage in the design process and the goals of the study. We then illustrate the utility of card sorting for Cartography by describing a closed sorting study we conducted on the ANCI INCITS 415-2006 emergency mapping symbol standard. Our approached helped us identify several barriers to using the symbol standard, including areas of conceptual overlap among the categories in the standard, potentially missing categories from the standard, and individual symbols in the standard that are consistently misclassified.

1. INTRODUCTION

Categorization is a fundamental way in which the humans make sense of their world, and therefore is a critical way in which knowledge is organized (MacEachren, 1995). The *card sorting* method is a knowledge elicitation technique designed to reveal the conceptual structures, or categorizations, of targeted individuals (Cooke, 1994). Card sorting requires participants to organize a set of instances, referred to as *cards* due to the original use of analog index cards, into internally-homogenous groups, or *categories*, according to similarity along an identified sorting principle, or *criterion* (Spencer, 2009). Depending on the card sorting variant, the contents of the cards may be different and the sorting criterion and categories may or may not be provided.

The potential of card sorting for both academic and practical Cartography is great. Card sorting is particularly promising for the design and organization of qualitative point symbol sets that primarily signify difference in kind, as the method works best for grouping cards on the same semantic level (Osgood et al., 1957) into nominal categories (McGeorge and Rugg, 1992) that have a single-level hierarchy (Wood and Wood, 2008). Card sorting can be used to categorize previously ungrouped map symbols, critique and refine existing symbol categorizations, identify missing or ambiguous symbols within the set, and suggest appropriate ways to represent the higher-level categorization within the individual symbols. Considering the automation of Cartography, card sorting also is a possible technique for injecting expert knowledge into a mapping system that automates the design of symbols and associated legend layouts (Rugg and McGeorge, 2005).

The research reported here is part of a broader line of work at the Penn State GeoVISTA Center on map symbology for emergency management and first response (Robinson et al., 2010). Specifically, this research is part of our work to develop a collaborative, yet distributed process for designing and sharing 'mission-specific' map symbol sets to support the range of missions within which government agencies use maps. Our focus in this paper, therefore, is on the practical application of card sorting by professionals who are seeking to solve a specific mapping problem, rather than its use as a scientific mode of inquiry to produce generalizable and repeatable results (although the latter is considered where appropriate and is the topic of other research at our Center).

The paper proceeds with five additional sections. In the following section, we review relevant work on card sorting in the fields of Psychology, Usability Engineering, and GIScience. We then outline a framework organizing variants of the card sorting method when applied to Cartography. In the fourth section, we introduce our case study and methodological approach. As a demonstration of the applicability of card sorting as a method, and as input to our larger research project on map symbology, we conducted a card sorting study with twenty Penn State undergraduates. Specifically, the study was designed to evaluate an existing emergency mapping qualitative point symbol standard designed by the Federal Geographic Data

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

Committee (FGDC) Homeland Security Working Group (ANSI INCITS 415-2006) to support emergency management and first response. We present and discuss the results of the card sorting study in the fifth section and offer concluding remarks and unanswered questions in the sixth and final section.

2. RELATED WORK

The card sorting method has its roots in Psychology. Eling et al. (2008) trace the method to the early 1920s, with the basic approach formalized as the *Wisconsin Card Sorting Test* (WCST) by the middle of the 20th century (Berg, 1946, Berg, 1948). The WCST is a repeat single-criterion sort (Rugg and McGeorge, 2005), requiring patients to sort the same set of cards multiple times, using a different criterion with each sort. The purpose of the WCST is to determine the patient's ability to think abstractly, extracting salient conditions from the set of instances and using variation in these conditions as criteria upon which to discriminate the instances into categories. Thus, the correct categorization for each sort is known *a priori* by the investigators and the patients are tested in their ability to reconstruct this categorization. Variants of the WCST still are used today in Neuropsychology to diagnose traumatic brain injuries (Stuss and Knight, 2002, Miller, 2007).

Usability engineers adapted card sorting for use as a method to extract expert knowledge from participants (see McDonald et al., 1986, for an early example). Here, the use of card sorting falls in line with a *user-centered design* (UCD) philosophy, which places an emphasis on obtaining input and feedback from the targeted end users of a software application early and often in the design/development process (Norman, 1988). The purpose of card sorting when used for usability evaluation is not to test the abilities of the participants, but instead to produce a "coherent design" that is specific to the information space under question (Nielsen and Sano, 1995: 182). Here, it is assumed that the targeted end users are the most authoritative experts on their own conceptual structures of an information space and that this knowledge can be elicited through card sorting to reveal or validate the optimal categorization.

Ensuring a "coherent design" is a primary consideration when compiling a map symbol set. Despite this, there are limited reported symbology design projects that leverage the card sorting method; a goal of the research reported here is to address this gap. One related set of studies examining the shape characteristics of star plot glyphs, a form of multivariate point symbol, is offered by Klippel et al. (2009a, 2009b). The goal of this work is to understand the perceptual and cognitive design considerations for all star plots rather than to design a map symbol set. Beyond symbology design, there are few reported interactive mapping and geovisualization software projects that use card sorting as a software usability technique, although several researchers have noted this potential of the card

sorting method (e.g., Robinson et al., 2005). A key exception is the work of Lloyd et al. (2008), who use card sorting to sort extant geovisualization techniques according to user tasks in the domain of crime mapping and analysis.

In the wider field of GIScience, there are two research thrusts in which card sorting is relevant: the cognitive grounding of qualitative spatial information theories and geographic ontologies. Researchers in Cognitive GIScience have developed qualitative formalisms that are used to abstract geographic detail and to create qualitatively equivalent classes to support efficient reasoning with spatial information. The cognitive adequacy of the proposed formalisms has to be evaluated before they are used to inform the design of spatial reasoning support systems. This evaluation commonly is achieved by using card sorting (a grouping paradigm) to assess whether both human conceptualizers and qualitative formalisms use the same abstraction principles. Particular research attention has been devoted to topological calculi such as the region connection calculus (Randell et al., 1992) and the 4/9 intersection models (Egenhofer and Franzosa, 1991). Other early research focused on static spatial relations (Mark and Egenhofer, 1994, Knauff et al., 1997, Renz, 2002), while more recent research has addressed geographic movement patterns (Klippel, 2009, Klippel and Li, 2009).

Ontologies assist in information sharing by aligning semantic similarities and differences across different knowledge domains (Gruber, 1993, Gruber, 1995); card sorting can be leveraged to establish the *is-part-of* relationship of ontologies. The use of ontologies for organizing geographic information is well established (see Schuurman, 2006, for an overview); examples applications of geographic ontologies include the integration of disparate geologic information (Brodaric et al., 2004) and land cover classification schemes of (Kavouras et al., 2005) as well as efforts to build a knowledge domain for geographic features (Dolbear and Hart, 2008, Ordnance Survey, 2010). Recent work has shown ontologies to be useful for compiling and aligning map symbols on various topographic map designs in the European Union (Hopfstock, 2007, Kent and Vujakovic, 2009). Therefore, card sorting may be particularly appropriate for developing geographic ontologies when sharing multiple symbols sets across different agencies (Brodaric et al., 2004). Such an application of card sorting allows for the integration and sharing of information across disparate systems and also minimizes the degree of designer subjectivity in ontological classifications and definitions, as with the testing of qualitative formalisms in Cognitive GIScience (Tomai and M.Kavouras, 2004).

3. CARD SORTING FOR MAP SYMBOLOGY

There are multiple methodological parameters that require close consideration when planning a card sorting study to support user-centered design. One overarching constraint is the current stage in the design process. When applied to

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

Cartography, card sorting can be used as both a pre-design method prior to constructing the symbol set or a post-design method following symbol set construction (Paul, 2008). This follows closely to the distinction between generative and evaluative usability studies introduced by Nielsen (2004)[1]. In *generative* card sorting, the investigators do not know the participants' classification of the information space and wish to reveal it. Conversely, in *evaluative* card sorting, the investigators have an existing classification and wish to assess it. Therefore, generative variants should be applied during pre-design and evaluative variants should be applied during post-design.

There are at least two important methodological parameters of card sorting that determine if the results will be generative or evaluative (or a blend of the two): the guidelines given to participants for sorting the cards into categories and the contents of the cards being sorted; each is

Table 1: A framework for determining the methodological parameters of a card sorting study when applied to inform design of a map symbol set.

considered in the following sub-sections. Table 1 provides an overview of the more common category-card combinations and the nature of the results likely to be obtained by each combination when applied for map symbol design. Once the goals of the card sorting study are established, this framework can be used to prescribe the experimental design. See Hannah (2005) for a review of other methodological parameters for the card sorting method.

3.1 Parameter #1: Guidelines Given to Participants for Sorting the Cards into Categories

Guidelines given to participants for sorting cards into categories generally fall into one of two types: open sorting and closed sorting (Spencer, 2009). In an *open sort*, participants are allowed to create their own categories as they complete the sort; they may or may not be constrained in the total number of categories they can create. Open sorts are generative in that they elucidate the sorting criterion and categories when both are unknown; therefore, it is more appropriate to administer open sorts during pre-design (Wood and Wood, 2008). Open sorts often produce a wider

variety of results than closed sorts. This can be a double-edged sword, as it makes category analysis and reconciliation more difficult, but also offers multiple, competing sorting criteria and category sets for follow up examination.

In a *closed sort*, conversely, participants are given the set of categories into which they must sort the cards. Closed sorts are evaluative in that they act to critique an existing classification constructed through either generative sorting or based on an *a priori* understanding of the information space; therefore, it is more appropriate to administer closed sorts during post-design (Tullis, 2007). Participants often are allowed to create an 'Other' or 'Discard' category in order to identify cards that are not on the same semantic level as the rest of the set (for both open and closed sorts), cards that do not align with the participant's chosen sort criterion (for open sorts), or cards that do not fit into any of the provided categories (for closed sorts) (Rugg and McGeorge, 2005).

We propose a third category: the guided sort. In a *guided sort*, the sorting criterion is provided, but the categories are not. Card sorting studies in Psychology avoid the use of directive, or guided, language to avoid biasing the results. Wood and Wood (2008) argue this is less necessary when the method is used functionally to develop a single product,

---

[1] While Nielsen (2004) introduced this distinction for usability methods generally, he only explicitly considered generative card sorting in his essay.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

as feedback is gathered more efficiently if the investigators adapt the sort to the problem at hand and emphasize known constraints. Guided sorts are different methodologically from open and closed sorts in that they are both generative (they produce previously unknown categories) and evaluative (they promote feedback on the provided sorting criterion); therefore, it is appropriate to administer them during pre- or post-design, depending upon the goals of the study.

3.2 Parameter #2: The Contents of the Cards Being Sorted

Rugg and McGeorge (2005) discuss three kinds of materials that can be sorted into categories during a card sorting study: objects, text (the original 'card' sort), and pictures. Many card sorting variants in Psychology require participants to categorize physical objects, such as colored strings or painted blocks (Eling et al., 2008). Object sorts are not practical for map symbol design due the geographic scale of the features of interests, although virtual environments such as Google Earth may offer one way in which this can be completed. Both text and picture sorts can assist in map symbol design, with text sorts used to evaluate feature definitions and picture sorts used to evaluate symbol designs; both are considered below.

In text sorts for map symbol design, feature names and/or definitions are placed on the cards for sorting. We generally recommend that both the feature name and definition are included in text sorts purposed for map symbol design. However, it is appropriate to include just the feature names when information on participants' interpretation of the terms is desired, such as when evaluating the labels used in the legend. Text sorts are more appropriate during pre-design, as information is needed on the sorting criterion and categories to inform symbol design, but the symbols are not yet available to sort. If the feature types are not yet known, alternative requirements assessment techniques should be administered. Additional feedback on the feature names and definitions can be collected by including a "Don't Understand" category.

In picture sorts, draft or established symbol designs are placed on the cards for sorting. There are two components of the symbol itself that can be evaluated during a picture sort: the graphical components that indicate the unique feature represented by the symbol (i.e., feature identification) and the graphical components that indicate the higher-level category in which the unique symbol belongs (i.e. category identification). Picture sorts are particularly good for evaluating the former (e.g., pictorial icons such as qualitative point symbols), as symbols that ambiguously represent their real-world referent will be sorted inconsistently across participants; this will be particularly evident if the same card was sorted consistently during pre-design text sorts. When the goal is to evaluate the sorting criterion or set of categories used for an established symbol set, it may be necessary to remove the graphical components of the symbols that indicate the high-

level category (e.g., the color scheme or frame shape used to indicate category) and to test only the graphical components that indicate the feature type. This is necessary because participants are more likely to validate the established categorization (whether it is appropriate or not) due to the presence of salient graphical clues. Finally, we recommend inclusion of both the feature definitions and the symbols on the cards for evaluative card sorting during post-design so that the definitions and symbols can be evaluated simultaneously.

4. METHODS

4.1 Case Study: The ANSI Map Symbol Standard for Emergency Management & First Response

The study reported here evaluates the ANSI INCITS 415-2006 symbol standard, referred to subsequently as the *ANSI standard*. The ANSI standard was designed by the Federal Geographic Data Committee (FGDC) Homeland Security Working Group to unify efforts in map design and map use for federal, state, and local missions related to emergency management and first response (Dymon, 2003, Dymon and Mbobi, 2005). The ANSI standard includes 198 unique qualitative point symbols categorized into four higher-level groups: (1) Incidents (features that symbolize the cause of action or source of disaster), (2) Natural Events (phenomena found in or created by naturally occurring conditions), (3) Operations (organizations, services, capabilities, or resources available during or implemented due to an emergency management situation), and (4) Infrastructures (the basic facilities, services, and installations needed for the functioning of a community or society).

Design of the symbols includes a central pictorial icon to designate the unique feature type and an outer frame that varies by frame shape to relate the symbol to the higher-level categorization. The incidents and natural events categories both are symbolized by the same frame shape, but differ in their use of whitespace; incidents are symbolized using a white background, with the pictorial icon marks in black, while natural events are symbolized using a black background, with the pictorial icon marks in white. The operations and infrastructures categories include four variants for each symbol to represent decreasing levels of operational status/increasing levels of damage, symbolized redundantly using frame pattern (the dashing used) and color. Finally, a unique feature type name and one sentence definition is provided for each symbol. Symbols and documentation for the ANSI INCITS 415-2006 standard are available at: http://www.fgdc.gov/HSWG/; examples from each category are provided in Table 2.

Despite the formal process for designing and disseminating the ANSI standard, it was not adopted widely by key mapmakers and map users at the Department of Homeland Security (DHS). A series of interviews was conducted with mapmakers and map users at seven DHS organizations to

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

	SYMBOL	TEST	NAME/DEFINITION
INCIDENTS			**residential fire:** a fire affecting a home or housing complex, resulting in partial or total destruction of the structure and/or bodily injury, smoke inhalation or death
			marine accident: a sudden, unexpected event involving a boat or ship resulting in vessel submerging, damage, bodily injury, death and/or the disruption of transportation service
NATURAL EVENTS			**drizzle:** some time called mist; very small, numerous, and uniformly dispersed water droplets that appear to float while following air currents; unlike fog droplets, drizzle fall to the ground
			landslide: a general term for a wide variety of processes and landforms involving the down slope movement under the force of gravity of masses of soil and rock material
OPERATIONS			**medical evacuation helicopter station:** the locus of an emergency helicopter landing pad, utilized to transport severely injured persons
			emergency water distribution center: a place where potable water is distributed to displaced persons or victims of war or disaster
INFRA-STRUCTURES			**toxic release inventory:** the location according to a publicly available database of chemical and other toxic waste releases
			traffic control point: the location of absolute signals controlled by an operator to regulate and maintain transportation flow

Table 2: Example symbols and definitions from the ANSI standard.

survey the degree to which the ANSI standard was adopted and to identify key barriers to its use (Robinson et al., 2010). The four-part symbol categorization was among the key barriers to use listed, as multiple agencies had their own internal structure that did not align exactly with the structure in the ANSI standard.

To understand this barrier to use, a card sorting study was conducted to illuminate general areas of confusion in the extant categorization as well as to identify individual symbols that are commonly misclassified. Because these goals were specific to a single map symbol set, our use of card sorting falls in line with the user-centered design approach taken by usability engineers focused on making better products, rather than the approach taken by psychologists or cognitive scientists focused on assessing and extending theory. Therefore, our experimental design, described in the following sub-section, was informed directly by the framework illustrated in Table 1.

4.2 Experimental Design

Following the framework in Table 1, we administered a closed sort that included the feature names, definitions, and symbols on the cards to support our goal of evaluating the ANSI standard. Additional methodological parameters requiring careful consideration include the study

participants, the number and contents of the cards, the testing environment, and the testing protocol; each is considered below.

The closed card sorting study was completed by twenty (n=20) Penn State undergraduates majoring in Geography. Participants were recruited from upper-level undergraduate Geography courses and emails over the Geography undergraduate listserv. An explanatory power study of the card sorting method conducted by Tullis and Wood (2004) indicated that a sample size of 15-30 participants explains 90-95% of the true information space structure, with diminishing returns in explanatory power as the sample size increases beyond 30; our sample size of 20 falls within this range of adequate explanatory power and efficient use of experimental resources. We felt it acceptable to test undergraduates because the primary goal of the study was to demonstrate the viability of card sorting as a method to be used in symbol set design and evaluation. We also believed that undergraduate majors in Geography would generate useful feedback about the ANSI standard, as they are more likely to have knowledge about maps and map symbol sets, although less likely to have knowledge about emergency response.

All 198 symbols were sorted during the exercise, which is within the accepted upper limit of 200 cards (Maurer and

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
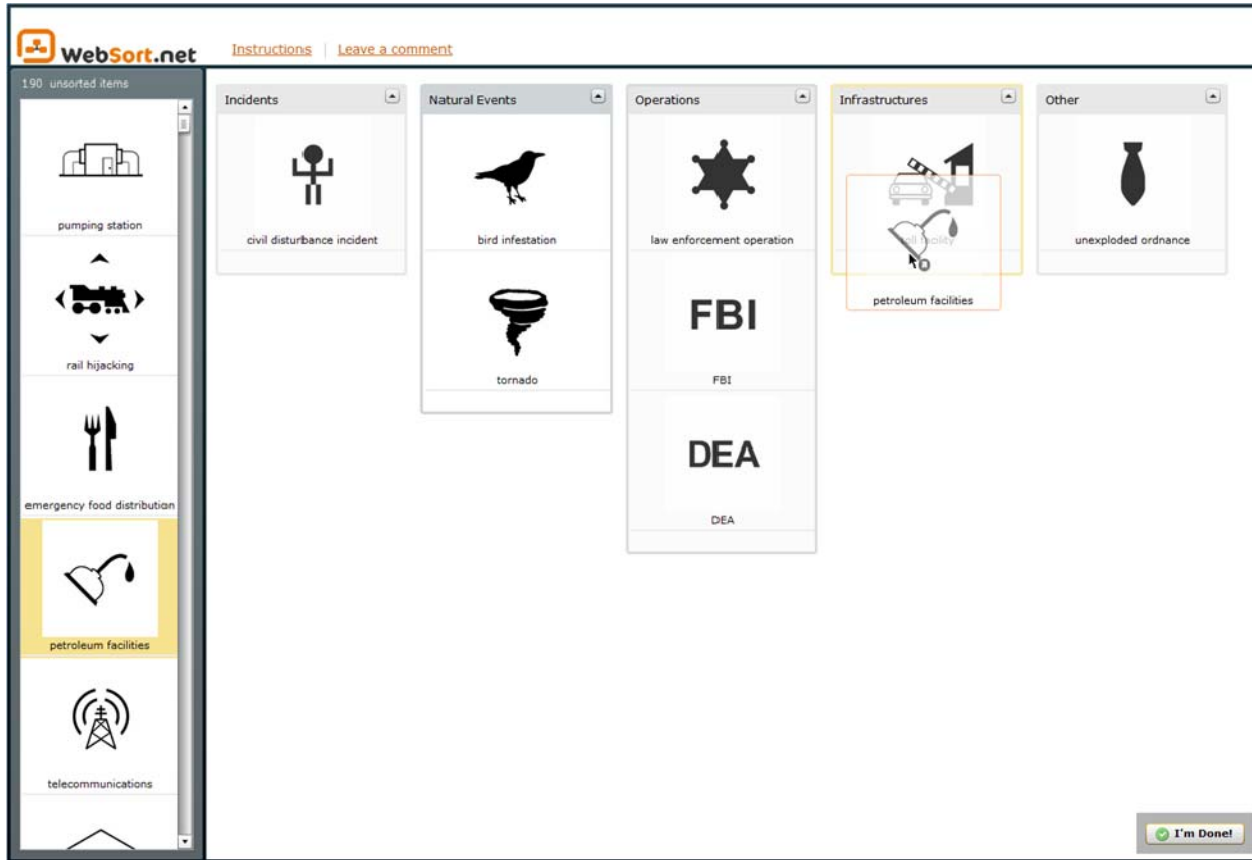November 15-19, 2010 Orlando, Florida

Figure 1: An example sort in the WebSort application.

Warfel, 2004, Nakhimovsky et al., 2006). Sorting 198 symbols was practical here because the closed sorting variant takes significantly less time to complete than the open sorting variant. Symbols were modified so that they did not include their outer frame shape indicating category in order to allow for evaluation of the established four-part categorization, the key barrier to using the ANSI standard identified by Robinson et al. (2010); example symbol modifications are shown in Table 2.

The card sorting study was administered using WebSort (http://www.websort.net), a web-based card sorting application. WebSort was chosen over other web-based sorting tools due to its superior ease in creating and analyzing sorting exercises (for a usability and utility comparison of other computerized card sorting applications to WebSort, see Chaparro et al., 2008). WebSort also supports open, guided, and closed sorts as well as text and picture sorts (or a combination of the two). Although the study could have been conducted in a distributed manner over the Internet, participants completed the sort in a controlled GeoVISTA Center lab to allow for the verbal communication of instructions and for a controlled setting while completing the sort (e.g., participants completed the sort in a set amount of time, they were not able to browse

other web pages, screen size and resolution remained constant, etc.).

Each card sorting session began by an investigator reading from a script that contained an overview of the study purpose, issues related to informed consent, and instructions for completing the sort; the latter portion of the script could be accessed by the participants for reference during sorting through the 'Instructions' link in WebSort. Users then were presented with the WebSort tool and given a brief demonstration of the drag-and-drop interface (illustrated in Figure 1). Five categories were provided for the closed sort: one for each of the four categories in the ANSI standard and a fifth 'Other' category for symbols perceived as falling outside the four-category structure. Symbols were presented in a random order and participants were not able to create additional categories. Definitions of the symbols were available as a tooltip upon mouse-over and definitions of the five categories were provided as a printed handout; all definitions were taken verbatim from the ANSI standard website. Participants were encouraged to refine their structures once completing an initial sort. Following completion of the sort, participants were asked to leave a comment about their general sorting strategy, any points of confusion during the sort, and specific symbols that were difficult to classify, particularly those placed in the 'Other' category. Participants were given 45 minutes to

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

complete the card sorting exercise; only one of the 20 participants was unable to finish in the allotted time, with most participants completing the sort in 20-30 minutes.

5. RESULTS AND DISCUSSION

The card sorting results were interpreted using both quantitative summary metrics and exploratory visual interpretation of associated information graphics (see Hannah, 2005, for a discussion on the relative advantages and disadvantages of either approach). In the following, quantitative summary metrics are used to identify key areas of potential concern with the overall ANSI standard categorization, while the exploratory visual interpretation of information graphics is used to enrich the understanding of these key problem areas by identifying specific symbols in the ANSI standard that are commonly misclassified.

5.1 Agreement

The most common calculation used to interpret sorting results is *agreement*, or the degree to which the participants' sorts are similar (Rugg and McGeorge, 2005). WebSort provides three measures of agreement: overall agreement by category (using the calculation from Spencer, 2009), card versus category (the percentage of sorts that placed a given symbol into a given category), and card versus card agreement (the percentage of sorts that placed two given symbols in the same category). Table 3 presents the overall agreement for each of the five categories included in the closed sort, as well as several additional summary statistics, to characterize the consistency across the twenty sorts.

The 'Infrastructures' (61%) and 'Natural Events' (56%) categories each exhibited an overall agreement over 50%,

standard due to the inclusion of a fifth 'Other' category. However, the average number of symbols placed in the 'Operations' category is notably smaller than its actual number of symbols (34.6 compared to 44, or a 21.4% decrease; no other category exhibited greater than a 12% decrease). A final measure is the number of unique symbols placed in each category (i.e., how many of the symbols in the ANSI standard that were placed into the category by at least one participant). All categories exhibited a considerable increase from the actual number of symbols in the category to the number of unique symbols placed in the category. 'Infrastructures' (55.6% increase) and 'Natural Events' (57.7% increase) exhibited the smallest increases in unique cards compared to their actual number of cards. Interestingly, while 'Operations' exhibited the greatest decrease in average cards per sort, it also exhibited the widest breadth of unique cards placed into the category (both by number, 129, and percentage increase, 193.2%). All three measures suggest that the 'Operations' category was particularly difficult to sort, resulting in little consistency across participants, and that the 'Infrastructures' and 'Natural Events' categories were the easiest to sort, resulting in the most consistency across participants.

5.2 Accuracy

The information presented in Table 3 considers only how participants performed against each other, and not how accurate they were in terms of recreating the actual four-category ANSI standard structure. To assess performance against the ANSI standard structure, a contingency table was derived from the card versus category agreement scores provided by the WebSort tool (Table 4). The 50% threshold was used to determine the category in which each symbol was placed in the majority of sorts, which can be considered as the 'mode' categorization. There were three

	AVG # PER CATEGORY	ACTUAL # IN CATEGORY	UNIQUE ITEMS	AGREEMENT
INCIDENTS	41.9	47	95	44%
INFRASTRUCTURES	77.1	81	126	61%
NATURAL EVENTS	23.0	26	41	56%
OPERATIONS	34.6	44	129	27%
OTHER	17.2	0	83	21%

Table 3: Sorting Agreement. The measures presented here indicate the consistency in performance across participants, with no relationship to the actual ANSI standard categorization.

the agreement threshold commonly used to assess similarity of sorting results (Paul, 2008). The 61% agreement in the 'Infrastructures' category is particularly positive given the large number of symbols it contains (81). There was unacceptably poor agreement in the 'Operations' category (27%), with slightly better agreement in the 'Incidents' category (44%). Expectedly, the average number of symbols per category in the twenty sorts was smaller than the actual number of symbols per category in the ANSI

symbols that were split evenly (50%/50%) into two categories across the twenty sorts (two 'Incident' symbols and one 'Natural Event' symbol); these symbols were placed into the 'No Majority' grouping despite being at 50% in one category, along with any symbols that were not placed into one of the categories in 50% or more of the sorts. Errors of commission run along the horizontal axis and errors of omission run along the vertical axis; errors of commission are not considered in the following discussion because of the asymmetry in the

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

SORT RESULTS (based on 50% agreement)	ANSI STANDARD				
	INCIDENTS	INFRA-STRUCTURES	NATURAL EVENTS	OPERATIONS	TOTAL
INCIDENTS	31	1	1	1	34
INFRASTRUCTURES	0	75	0	12	87
NATURAL EVENTS	2	0	24	0	26
OPERATIONS	0	1	0	28	29
OTHER	7	0	0	0	7
NO MAJORITY	7	4	1	3	15
TOTAL	47	81	26	44	198
% CORRECT	66.0%	92.6%	92.3%	63.6%	79.8%

Table 4: Sorting Accuracy. A contingency table was derived from the card versus category agreement scores to compare the actual ANSI standard categorization to the mode categorization generated by the twenty sorts.

contingency table due to the inclusion of an 'Other' category during the sort and the lack of majority in 15 of the 198 symbols.

Overall, nearly 80% of the symbols were sorted correctly a majority of the time, a quite large percentage considering the size and complexity of the information space. As with the across participant measures in Table 3, the 'Infrastructures' (92.6%) and 'Natural Events' (92.3%) categories exhibited the highest accuracies. The 'Operations' category again was the most problematic, with only 63.6% of the symbols categorized correctly a majority of the time. Interestingly, most of the misclassified symbols (12 of 16) were placed into 'Infrastructures' category, suggesting that there may be a subset of symbols that should be moved from 'Operations' to 'Infrastructures' or that conceptually span the two categories (see Section 5.3 for further discussion). For the 'Incidents' category, the accuracy score (66.0%) was relatively lower when compared to the other categories than its agreement score. Most of the misclassified 'Incidents' symbols (14 of 16) were placed into the 'Other' category or did not receive a majority, suggesting that a portion of the 'Incidents' category may belong in its own independent category or in a different category (again, see Section 5.3 for further discussion).

5.3 Bivariate Matrix

The primary information graphic leveraged for exploratory interpretation of the closed sorting results was a bivariate matrix of agreement scores. Bivariate matrices were produced for both card versus card agreement (Figure 2, large section on the left/in the center) and card versus category agreement (Figure 2, thin section on along right border). The order of the symbols is determined using the average linkage hierarchical clustering algorithm provided

by WebSort, which builds a nested hierarchy based on the agreement scores (O'Sullivan and Unwin, 2003); thus, symbols that were more likely to be found in the same category during the twenty sorts are more likely to be close to each other in the matrix. A high resolution, color version of the information graphic with symbol labels and embedded agreement scores is available for download at: http://www.geovista.psu.edu/Symbology/CardSorting/. It is important to note that additional information graphics were used during the exploratory visual analysis, but are not presented her due to space limitations; these additional graphics are posted at the aforementioned site and include a dendogram showing the results of the hierarchical clustering and a spatialization plot showing the results of multidimensional scaling.

The axes of the bivariate scheme shown in Figure 2 are same as the metrics included in Tables 3 and 4: agreement (either under 50% or at 50% or above) and accuracy (either correct or incorrect pairings). Cells colored in either black (>=50% agreement and correct) or white (<50% and incorrect) are appropriate pairings of agreement and accuracy, reflecting areas where the sorts conformed to the ANSI standard categorization. Cells colored in either orange (>=50% agreement, but incorrect) and purple (<50% agreement, but correct) indicate areas where the sorts do not conform to the ANSI standard categorization. The general locations of the four ANSI standard categories are labeled along the card versus card agreement matrix; each ANSI standard category is considered below, with specific subsets of symbols that are discussed labeled by letter in the matrix.

The 'Operations' category is identified in Tables 3 and 4 as the worst in both agreement and accuracy. Viewing the bivariate matrix, it appears as though participants handled

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

two large subsets of the 'Operations' symbols differently. Subset A consists of 15-20 symbols that were correctly classified in a majority of sorts, suggesting that participants clearly identified this set of symbols as 'Operations'. Subset A is composed of symbols with the word 'emergency' in their name (e.g., 'emergency operations center', 'emergency staging areas', 'emergency teams'), medical related symbols (e.g., 'ambulance', 'hospital ship', 'medical evacuation helicopter station'), and the set of six sensor-related symbols. Thus, participants took the 'Operations' category to mean symbols either related to detecting an emergency or responding to one.

Subset B is composed of 10-12 'Operations' symbols that exhibit an extremely high degree of card versus card agreement (most symbol pairs are at 85% or higher, with several pairs receiving 100% agreement), meaning that

most participant saw this group of symbols as a homogenous unit. Subset B is composed of symbols identifying DHS agencies (e.g., 'FBI', 'secret service', 'US Coast Guard') or municipal agencies (e.g., police). Despite the high level of card versus card agreement, most of the symbols correctly marked as 'Operations' have only a 50% card versus category agreement. The large amounts of purple and orange extending from this subset indicate that many of the participants chose to locate this entire group of symbols in the 'Infrastructures' category instead, likely leading to the poor overall agreement and accuracy reported for the 'Operations' category in Section 5.1. This confusion may be a result of testing undergraduates unfamiliar with emergency operations, but also may be explain partly by ambiguity in the 'Operations' definition provided at the ANSI standard website. Further, several participant comments indicated that this subgroup may be
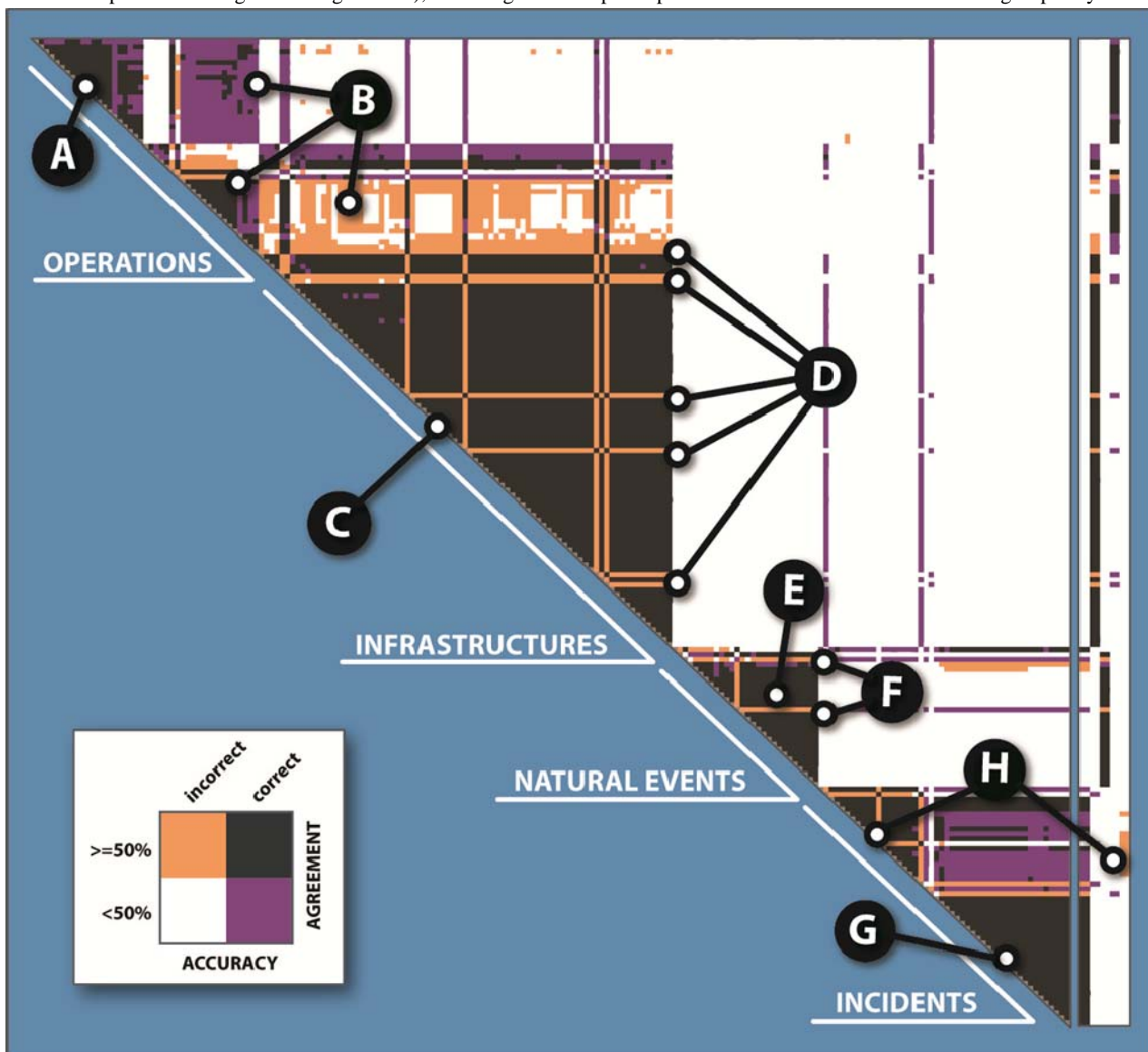

Figure 2: A bivariate matrix showing agreement versus accuracy for card versus card (left) and card versus category (right) pairings.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

best represented as a separate category called "Organizations".

The 'Infrastructures' category is identified in Tables 3 and 4 as the strongest in both agreement and accuracy. This is reflected congruently in Subset C, a group of 75-80 symbols that are sorted correctly into the 'Infrastructures' category with a high level of agreement among participants; nearly all card by category and card by card agreement scores are above 75%, with many above 90%. Aside from the common conflation with Subset B, as described above, there is a small subset of problematic 'Operations' symbols sprinkled through Subset C. These symbols, marked by the non-contiguous Subset D, include 'other water supply location', 'fire hydrant', 'fire station', 'medical facilities outpatient', 'morgue', 'health department facility', 'prison', 'hospital', and 'pharmacies'. The sorting revealed that the permanent installation of these features in the built environment (i.e., as infrastructure) is more salient than their relationships to emergency operations; thus, the symbols contained in Subset D should be moved from 'Operations' to 'Infrastructures' or the function in operations should be more clearly indicated in the symbol description.

The 'Natural Events' category also is characterized with high levels of agreement and accuracy in Tables 3 and 4. As expected, the associated Subset E contains 25-30 symbols that are correctly categorized as 'Natural Events' and that exhibit a high degree of card versus card and card versus category agreement. There are two important symbol exceptions to this otherwise homogenous subset worth discussing, marked as the non-contiguous Subset F. Participants commonly sorted the 'Incidents' symbols 'smoke' and 'wildfire' into the 'Natural Events' categories. Both feature types can be caused naturally and anthropogenically, a point of confusion for participants. It may be useful to include two variants of the symbol to differentiate the two types of occurrences. As one participant noted in a comment, this may be the case for the 'hotspot' symbol as well.

Like the 'Operations' category, symbols in the 'Incidents' category subdivide into two primary groupings in the matrix. Subset G contains 25-30 symbols that were categorized correctly as 'Incidents' by nearly all participants; almost all of these symbols have a card versus category agreement of 95% or more and card versus card agreement of 90% or more. This subset is composed of symbols for air-, civil-, fire-, marine-, rail-, vehicle-related occurrences, all of which represent unnatural events, rather than objects. In contrast, Subset H contains 18-20 symbols that overall are categorized as 'Incidents' correctly, but have much greater card versus card agreement scores than card versus category agreement scores. Broadly described, this subset contains symbols representing materials that cause critical incidents, rather than symbols representing the critical incident itself (as with Subset G); examples include 'bomb', 'explosive', 'chemical agent', 'toxic gas', and 'flammable gas/liquid/solid'. Interestingly, a majority of participants placed 10 of these symbols into the 'Other'

category, distinguishing this subset as an additional category; these were the only 10 symbols in the sorting exercise that were placed in the 'Other' category by a majority of participants. Thus, it may be appropriate to create an additional category to distinguish 'Incidents' from the objects causing these incidents. Names suggested in the comments for this additional category include "Potential Incidents" and "Potentially Dangerous".

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have shown the value of applying the card sorting method to evaluate and refine map symbol sets. We also have introduced a framework that prescribes two key methodological parameters—guidelines for sorting cards into categories and the contents of the cards being sorted—according to the stage of the symbology design process and the goals of a symbology study. In a case study application, we presented a closed sort of symbols and feature definitions using the ANSI INCITS 415-2006 emergency mapping symbol set. This study identified key problems in the 'Incidents' and 'Operations' categories of the ANSI standard, as well as numerous individual symbols that are consistently misclassified.

In future work on emergency mapping, we plan on integrating card sorting into a distributed Delphi process aimed at generating and sharing 'in-house' standards specific to each mission area within the Department of Homeland Security. There is still much more to understand about the use of card sorting for map symbol design. Unanswered questions include:

- How can the method be used to generate multi-level categorizations?
- What types of statistical and visual interpretation methods should be applied for each card sorting study variant?
- When should the sorting be conducted collaboratively and when should participants be allowed to discuss their results to develop a final categorization?
- How do sorting results relate to fuzzy sets and prototype theory and how can the results be used as measures of relevancy or certainty?
- How can sorting be used to build map symbol ontologies and allow for role-based symbology translation and visualization?

Supporting mapmakers and map users with effective and understandable symbology will require answers to these questions (among others). The card sorting approach, like the one we describe here, is one way to begin to understand and rectify symbology challenges that currently impede collaboration and situational awareness in a wide range of geographic application domains.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

REFERENCES

Berg, E., 1946. An objective technique for the measurement of flexibility in thinking. MS Thesis, Psychology Department, University of Wisconsin, Madison, WI, USA.

Berg, E. A., 1948. A simple objective for measuring flexibility in thinking. *Journal of General Psychology*, 39, pp. 15-22.

Brodaric, B., M. Gahegan, & R. Harrap, 2004. The art and science of mapping: Computing geological categories from field data. *Computers & Geosciences,* 30(7), pp. 719-740.

Chaparro, B. S., V. D. Hinkle, & S. K. Riley, 2008. The usability of computerized card sorting: A comparison of three applications by researchers and end users. *Journal of Usability Studies,* 4(1), pp. 31-48.

Cooke, N. J., 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies,* 41(6), pp. 801-849.

Dolbear, C., & G. Hart, 2008. Ontological bridge building - using ontologies to merge spatial datasets. *AAAI Spring Symposium on Semantic Scientific Knowledge Integration, AAAI/SSS Workshop.* 26-28 March, Stanford, CA, USA.

Dymon, U. J., 2003. An analysis of emergency map symbology. *International Journal of Emergency Management,* 1(3), pp. 227-237.

Dymon, U. J., & E. K. Mbobi, 2005. Preparing an ANSI standard for emergency and hazard mapping symbology. *International Cartographic Conference.* 9-16 July, A Coruña, Spain.

Egenhofer, M. J., & R. D. Franzosa, 1991. Point-set topological spatial relations. *International Journal of Geographical Information Systems,* 5(2), pp. 161-174.

Eling, P., K. Derckx, & R. Maes, 2008. On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition,* 67(3), pp. 247-253.

Gruber, T. R., 1993. A translation approach to portable ontologies. *Knowledge Acquisition,* 5(2), pp. 199-220.

Gruber, T. R., 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies,* 43(4-5), pp. 907-928.

Hannah, S., 2005. Sorting out card sorting: Comparing methods for information architects, usability specialists, and other practitioners. MS Thesis, Applied Information Management, University of Oregon, Portland, OR, USA.

Hopfstock, A., 2007. A comparative study of European topographic maps at scale 1:250,000 (with particular emphasis on motorways). *International Cartographic Conference.* 4-10 August, Moscow, Russa.

Kavouras, M., M. Kokla, & E. Tomai, 2005. Comparing categories among geographic ontologies. *Computers & Geosciences,* 31(2), pp. 145-154.

Kent, A., & P. Vujakovic, 2009. Stylistic diversity in European state 1:50000 topographic maps. *The Cartographic Journal,* 46(3), 179-213.

Klippel, A., 2009. Topologically characterized movement patterns: A cognitive assessment. *Spatial Cognition and Computation,* 9(4), pp. 233-261.

Klippel, A., F. Hardisty, R. Li, & C. Weaver, 2009a. Color enhanced star plot glyphs - Can salient shape characteristics be overcome? *Cartographica,* 44(3), 217-232.

Klippel, A., F. Hardisty, & C. Weaver, 2009b. Star Plots: How shape characteristics influence classification tasks. *Cartography and Geographic Information Science,* 36(2), 149-163.

Klippel, A., & R. Li, 2009. The endpoint hypothesis: A topological-cognitive assessment of geographic scale movement patterns. In: K.S. Hornsby., C. Claramunt, M. Denis, & G. Ligozat (Eds.) *9th International Conference on Spatial Information Theory (COSIT).* Springer, Aber Wrac'h, France, pp. 177-194.

Knauff, M., R. Rauh, & J. Renz., 1997. A cognitive assessment of topological spatial relations: Results from an empirical investigation. In: S.C. Frank & A.U. Frank. (Eds.) *Spatial information theory: A theoretical basis for GIS.* Springer: Berlin, Germany.

Lloyd, D., J. Dykes, & R. Radburn, 2008. Mediating geovisualization to potential users and prototyping a geovisualization application. In: D. Lambrick (Ed.) *GIS Research UK.* 2-3 April, Machester, UK.

MacEachren, A. M., 1995. *How maps work.* The Guilford Press, New York, NY, USA.

Mark, D. M., & M. J. Egenhofer, 1994. Calibrating the meanings of spatial predicates from natural language: Line-region relations. In: T.C. Waugh & R.G. Healey (Eds.) *Advances in GIS Research, 6th International Symposium on Spatial Data Handling.* Edinburgh, Scotland, pp. 538-553.

Maurer, D., & T. Warfel, 2004. "Card sorting: A definitive guide". http://www.boxesandarrows.com/view/card_sorting_ a_definitive_guide (accessed 10 Sept. 2010).

McDonald, J. E., D. W., Dearholt, K. R. Paap, & R. W. Schvaneveldt, 1986. A formal interface design methodology based on user knowledge. *Human Factors in*

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida

Computing Systems (CHI). ACM, Boston, MA, USA, pp 285-290.

McGeorge, P., & G. Rugg, 1992. The uses of 'contrived' knowledge elicitation techniques. *Expert Systems,* 9(3), 149-154.

Miller, B., 2007. *The human frontal lobes: Functions and disorders.* Guilford, New York, NY, USA.

Nakhimovsky, Y., R. Schusteritsch, & K. Rodden, 2006. Scaling the card sort method to over 500 items: Restructuring the Google AdWords Help Center. *Human Factors in Computing Systems (CHI).* ACM, Montreal, Quebec, pp. 183-188.

Nielsen, J., 2004. "Card Sorting: How many users to test." http://www.useit.com/alertbox/20040719.html (accessed 10 Sept. 2010).

Nielsen, J., & D. Sano, 1995. SunWeb: User interface design for Sun Microsystem's internal Web. *Computer Networks and ISDN System,* 28(1-2), 179-188.

Norman, D. A., 1988. *The design of everyday things.* Basic Books, New York, NY, USA.

O'Sullivan, D., & D. J., Unwin, 2003. *Geographic information analysis.* John Wiley & Sons, Hoboken, NJ, USA.

Osgood, C. E., G. J. Suci, & P. H. Tannenbaum, 1957. *The measure of meaning*. University of Illinois Press, Urbana, IL, USA.

Paul, C. L., 2008. A modified Delphi approach to a new card sorting methodology. *Journal of Usability Studies,* 4(1), 7-30.

Randell, D. A., Z. Cui, & A. G. Cohn, 1992. A spatial logic based on regions and connections. *3rd International Conference on Knowledge Representation and Reasoning.* Morgan Kaufmann, San Francisco, CA, USA, pp. 165-176.

Renz, J., 2002. Qualitative spatial reasoning with topological information. *Lecture Notes in Artificial Intelligence.* Springer Berlin, Germany.

Robinson, A. C., J Chen, E. J. Lengerich, H. G. Meyer, & A. M. MacEachren, 2005. Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science,* 32(4), 243-255.

Robinson, A. C., R. E. Roth, & A. M. MacEachren, 2010. Challenges for map symbol standardization. *7th International Conference on Information Systems for Crisis Response and Management.* 4 May, Seattle, WA.

Rugg, G., & P. McGeorge, 2005. The sorting techniques: A tutorial paper on card sorts, picture sorts, and item sorts. *Expert Systems,* 22(3), 94-107.

Schuurman, N, 2006. Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers,* 96(4), 726-739.

Spencer, D., 2009. *Card Sorting: Designing usable categories,* Rosenfield Media, Brooklyn, NY, USA.

Stuss, D. T., & R. T. Knight, 2002. *Principles of frontal lobe function.* Oxford University Press, New York, NY, USA.

Ordnance Survey, 2010. "Ordnance Survey Ontologies". http://www.ordnancesurvey.co.uk/oswebsite/ontology/ (accessed 10 Sept. 2010).

Tomai, E., & M. Kavouras, 2004. From "Onto-GeoNoesis" to "Onto-Genesis": The Design of Geographic Ontologies. *GeoInformatica,* 8(3), 285-302.

Tullis, T., & L. Wood, 2004. How many users are enough for a card-sorting study? *Usability Professionals Association.* 7-11 June, Minneapolis, MN.

Tullis, T. S., 2007. Using closed card-sorting to evaluate information architectures. *Usability Professionals Association.* 11-15 June, Austin, TX.

Wood, J. R., & L. E. Wood, 2008. Card Sorting: Current practices and beyond. *Journal of Usability Studies,* 4(1), 1-6.

A special joint symposium of ISPRS Technical Commission IV & AutoCarto
in conjunction with
ASPRS/CaGIS 2010 Fall Specialty Conference
November 15-19, 2010 Orlando, Florida