R. Scott Hanrath. A Usability Study of a Tool for Contributor-supplied Metadata Creation: the Use of Metadata Element Definitions and Examples in Online Help. A Master's paper for the M.S. in I.S. degree. November, 2002. 30 pages. Advisor: Jane Greenberg

This paper describes a usability study of Botanical Pride, a contributor-supplied metadata creation tool which allows botany enthusiasts to create metadata for images of botanical specimens. Two versions of the interface were tested: a Definitions Only version that included definitions of metadata elements and a Definitions+Examples version that included definitions of metadata elements and one or more examples of metadata element values. Participants used each version of the interface to create a metadata record for a different selected image. Testing was conducted via the World Wide Web. Usability was measured by responses to a satisfaction questionnaire and by mean task completion times. Recommendations are made for the use of examples in the help for Botanical Pride. Implications for the design of future studies of metadata creation tools for use by non-metadata experts are considered.

Headings:

Metadata

Author-generated Metadata

Usability

A USABILITY STUDY OF A TOOL FOR CONTRIBUTOR-SUPPLIED METDATA CREATION: THE USE OF METADATA ELEMENT DEFINITIONS AND EXAMPLES IN ONLINE HELP

By
R. Scott Hanrath

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science

Chapel Hill, North Carolina

November, 2002

Approved by:

_____
Advisor

Table of Contents

Introduction

Metadata is increasingly being recognized as an essential component of information systems. Defined as structured data about data, metadata describes a resource or object in order to support discovery, access, manipulation, and often other functions. Burnett, Ng, and Park (1999) note that the emergence of the Word-Wide Web has fueled interest in metadata as the concerns of the library cataloging and data management communities, two communities with long standing interests in metadata, move closer together. As a result, a number of efforts have concentrated on improving metadata on the Web, where the lack of metadata and metadata standards can make the discovery and evaluation of information a difficult, frustrating experience. For example, the Dublin Core Metadata Initiative (DCMI) has developed a widely adopted 15 element set for simple resource description (DMCI, 1999) and the Word-Wide Web Consortium's proposed Resource Description Framework (RDF) and RDF Schema have provided a data model and syntax for expressing metadata in support of what is envisioned as the Semantic Web (Miller, 1998).

As information systems move from the domain of the technologically-savvy to tools and resources available to more general populations, the usability of such systems takes on an increased importance. This is especially true of the Web, which by its nature is more broadly accessible than most previous information systems. The research contribution to metadata schemas and standards for the Web has been considerable and the library community has a long history of inter-indexing consistency studies, which measure the degree to which multiple indexers agree when assigning terms to represent document subjects (Markey, 1984). Less attention has been paid to the usability of tools

for metadata creation, especially those intended to support non-specialists (as opposed to information specialists) in a Web environment.

It is reasonable to expect that for metadata to become commonplace and useful on the Web, it must be harnessed to usable systems. Further, end users need not just usable systems that effectively employ metadata to facilitate resource discovery and evaluation, but usable systems that facilitate the creation of metadata. Whether for posting an item for sale in an online auction or adding an object to a public digital library collection, usable systems that facilitate the creation of metadata by non-specialist author/creators are required.

Studies of web usability have identified "best practices" for aspects such as page, content, and site design (e.g., facilitating scanning through clear headings and bulleted lists and improving on-screen readability by using less text and an inverted pyramid writing style) (Nielsen, 2000). Similar research is needed to help improve interfaces for the creation of author-generated metadata by identifying how different factors affect their usability. Research in this area has the potential to make a valuable contribution to areas from e-commerce to digital libraries, where an array of people not trained as information professionals are engaged in metadata creation.

Literature Review

Metadata Creation

Duval, Hodgins, Sutton, and Weibel (2002) note that there is little agreement about how metadata should be integrated into information systems. Along with the issues of metadata standards (as addressed by, e.g., the Dublin Core), representation (as

addressed by, e.g., the Resource Description Framework), exchange (as addressed by, e.g., the Open Archives Initiative) is the question of metadata creation. Thomas and Griffin (1999) note that the problem of metadata creation has not received the same attention as other metadata issues. Given the potential importance of resource metadata, how should it be created? And who should be responsible for creating it?

Thomas and Griffin (1999) argue that metadata framework efforts "misjudge the degree to which schemas will be implemented" and that the assumption that resource creators will be responsible for metadata creation is flawed. They pose the problem in economic terms, arguing that in the case of business organizations metadata creation can be perceived as expensive for the value it provides; they suggest that commercial indexing services, not content creators, have the strongest financial incentive to create metadata (Thomas & Griffin, 1999). Milstead and Feldman (1999) question not the incentive of resource creators to create metadata, but their ability: "how do we get millions of non-information professionals to understand the importance of cataloging to a certain level and standard when even professionals don't always agree?"

Still, most agree that it is highly likely that resource creators or contributors will be in some way responsible for metadata creation. A study by Greenberg, Pattuelli, Parsia, and Robertson (2001) tested the hypothesis that "given basic guidance through a simple and intelligible Web form, resource authors can create professional quality metadata." The hypothesis was confirmed by the study, in which all metadata created by resource authors was found to acceptable pending some revisions. Greenberg et al., (2001) suggest that better textual guidance for authors might further increase the quality of author-supplied metadata. Similarly, Duval, Hodgins, Sutton, and Weibel (2002)

suggest that sophisticated facilities for metadata creation can simplify the process for resource authors. Such applications can make it easier to combine creator-supplied and automatically generated metadata, and thus both improve the quality and decrease the cost of metadata (Duval et al., 2002).

Defining Usability

Usability, then, appears to be an important factor in encouraging metadata creation. Usability as a concept is implicitly related to the process of User-Centered Design (UCD), a process that has been developed in such overlapping areas as human factors engineering, ergonomics, and usability engineering (Rubin, 1994). UCD seeks to put humans ("users"), rather than systems at the center of the information system design process. Nielsen and Mack (1994) offer a loose definition of usability as "a fairly broad concept that basically refers to how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use" (p. 3). A widely accepted definition of usability is that of ISO/DIS 9241-11: "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (European Usability Support Centres, 2001).

Justifying usability

Why is usability important? Rubin (1994) cites five reasons why technical products are difficult to use, all of which stem from a lack of focused attention to usability. These are: (1)focus on system/machine, not user; (2)typical users are becoming less sophisticated; (3)design of usable systems is difficult, yet often treated as a

afterthought in design process; (4)participants in the design process are often specialists whose efforts are often not integrated; (5)design and implementation of user interfaces require distinct skill sets, yet those trained in implementation often are responsible for both tasks (pp. 4-10). All of these are potential problems for contributor-supplied metadata creation tools.

Aspects of usability

Three widely used aspects of usability are those cited in the ISO definition: effectiveness, efficiency, and satisfaction (see Figure 1). Frøkjær, Hertzum, and Hornbæk (2000) clarify effectiveness as "the accuracy and completeness with which users achieve certain goals," efficiency as "the relation between [effectiveness] and the resources expended in achieving it," and satisfaction as "the user's comfort with and positive attitudes toward the use of the system" (p. 345).

Aspect	**Definition (adapted from Frøkjær et al., 2000)**
Effectiveness	Accuracy and completeness with which users achieve certain goals
Efficiency	Relationship between effectiveness and the resources expended in achieving it
Satisfaction	User's comfort with and positive attitudes toward the use of the system

Figure 1. Aspects of Usability

Both Rubin (1994) and Nielsen and Mack (1994) additionally include learnability, a factor, not wholly accounted for in the ISO aspects, where efficiency may include experienced as well as novice users.

Frøkjær et al. (2000) cite evidence from a study that effectiveness, efficiency, and satisfaction are at best weakly correlated statistically. The authors of the study note that most experimental studies from a survey of ACM CHI Proceedings ignore at least one of

the three usability aspects and suggest that such studies either assume an unverified correlation between aspects, or risk ignoring important aspects (Frøkjær et al., 2000, p. 351).

Nielsen (2000) approaches usability on the web from the standpoint of guidelines for various aspects of websites, organizing his discussion of usability around best practices for page design, content design, and site design. Zhang and von Dran (2000) note that this is a common approach in web usability literature, where design checklists similar to Nielsen's are more common than theoretical models.

Zhang and von Dran (2000) propose such a theoretical model for the design and evaluation of website factors. Zhang and Dran (2000) base their model on Herzberg's motivational management theory. Under this model, website features fall into two categories, those which are "hygiene" features, the absence of which leads to user dissatisfaction and those which are "motivator" features, the presence of which contributes to user satisfaction. Hygiene factors provide the necessary, but not sufficient conditions for user satisfaction. Zhang and Dran (2000) offer their model as a framework for explaining previous research that showed users to prefer (i.e., be more satisfied with) systems other than those which proved to be the most effective (p. 1247).

Many authors note the importance of taking a system's application domain into account in assessing which usability aspects may be most important. Frøkjær et al. (2000) note that while effectiveness and efficiency may not be correlated, effectiveness may be more important than efficiency for complex tasks, while the opposite may hold true for routine tasks (p. 345). Zhang and von Dran (2000) caution that the same feature that functions as a hygiene factor for website in one domain, may function as a motivator

factor in another (e.g., visual appeal may be a hygiene factor for an entertainment website, but a motivator for an educational website) (p. 1263). Nielsen (1994) notes that in conducting heuristic evaluations, the more a system is focused on a particular domain, the more important it is supplement open-ended evaluations with specific scenarios.

Usability of Metadata Creation Tools

A literature search revealed little research on the usability of metadata creation tools. A number of studies have explored the display, rather than creation of metadata. Studies involving information visualization and metadata have examined the effect of using different means to display metadata to users. For example, Kumar, Furuta, and Allen (1998) explored different means of incorporating metadata into event timelines in a hypertext environment, but provided no usability data. Fraser and Gluck (1999) examined how end-users used metadata to evaluate the relevance of geospatial information resources in a exploratory study that observed subjects conducting searches using various metadata record display formats. They found that format, length, order of attributes, geographic resolution, and time were important factors in the usability of geospatial metadata usability and suggest that metadata standards bodies have much to gain in terms of their standard's acceptance by focusing on usability concerns.

Harmes (2001) suggests that the inclusion of detailed help in the form of metadata element descriptions and examples are among the important features of a usable form for capturing author-generated metadata, a suggestion also made by Greenberg et al. (2001). The author of this paper previously conducted an exploratory study that analyzed the textual help features of four actively used metadata creation tools spanning different communities (from open source software to education materials) and displaying a range

of complexity (from 4 to 24 elements) (Hanrath, 2002). The previous study found that definitions of metadata elements (i.e., a short statement describing an element) and examples of metadata element values were commonly employed in online help and best practices documents for metadata creation tools. The kind and depth of help provided by definitions and examples, however, varied greatly across the tools. For example, while definitions were the most commonly observed type of help, they were not consistently present, suggesting the lack of a de-facto minimum standard. Examples were also common, but were used in at least three distinct ways: Syntax examples (e.g., "Last name, First name"), Generic examples (e.g., "poet, author, or painter"), and Instance examples (e.g., "Walt Whitman"). The choice of definitions and examples appeared to be related to the robustness of the metadata scheme (e.g., the number of elements and semantic relationships between elements), the intended domain or community (including the relative incentive for the user to create the metadata), and the complexity of the interface (e.g., the use of "implicit" element value examples in pull-down menus for controlled vocabularies).

Nielsen (2000) argues that usability is of greater importance to websites than to other technical products, such as appliances or desktop software applications, due to divergent business models. For traditional software applications the user often pays first, and experiences usability after the fact. On the web, Nielsen (2000) claims, this process is often reversed, with users having the opportunity to experience—and evaluate—the usability of a product before committing to payment. A similar argument can be made about the usability of interfaces for author- or contributor-supplied metadata: systems lacking in usability may discourage quality metadata creation, particularly in cases where

users do not have personal incentives for providing metadata. The usability of metadata creation tools as measured by satisfaction deserves special attention by researchers. In order to encourage metadata creation, we must provide users with tools which they are comfortable and towards which they display positive attitudes.

Objectives of the Study

This paper seeks to asses the importance of two help features for contributor-supplied metadata creation tools: the inclusion of definitions of each metadata element in the available help and the inclusion of domain-specific examples for each metadata element in the available help. The purpose of the study was two-fold. First, the study was intended to collect usability data to improve a particular interface, Botanical Pride. Second, it was intended to serve as an exploratory study of the relative importance of the use of metadata elements descriptions and examples in the help of contributor-supplied metadata creation tools in general.

Method

Because little is known about the usability of interfaces for capturing author-generated metadata, user testing, rather than heuristic evaluation, was selected as the primary research method.

This study examined the impact of two different help features on the usability of a form for capturing author-generated metadata. The interfaces tested were based on the Botanical Pride interface developed by Harmes (2001). The Botanical Pride interface was designed to allow the general public to contribute images of botanical specimens to an online database along with corresponding metadata records. The interface form

includes six Dublin Core Metadata elements: Title, Description, Subject, Coverage, Date, and Creator. Dublin Core was selected for its general applicability and simplicity. The labels for three elements were altered: Name was used for Title, Keyword for Subject, and Geographic Location for Coverage.

Interfaces Tested

The study tested two versions of the Botanical Pride-based interface. The Definitions Only Version included metadata element labels and definitions for each element. The Defintions+Examples Version included metadata element labels, definitions for each element, and one or more domain-specific examples for each element (see Appendix A).

Study Participants

Participants were recruited from a state botanical gardens, a university-affiliated arboretum, and an undergraduate student botany club in order to provide a study group with relevant domain experience. Participants were solicited via email and participated on a volunteer basis. Participants were screened to ensure the exclusion of participants with significant metadata or cataloging experience.

Measures

Because providing incentives for contributors to provide metadata is seen as a key concern, the usability of the two interfaces was assessed by measures of satisfaction and efficiency.

Items selected from the Questionnaire for User Interface Satisfaction (QUIS) were administered following the testing of each interface. Research has demonstrated that QUIS has good reliability and external validity in evaluating user satisfaction with human-computer interfaces (Chin, Diel, & Norman, 1988). QUIS was developed to test more complex computer systems than the simple interfaces under consideration in this study so not all the items on QUIS were applicable. The study reported here used the six items from the "Overall Reactions to the Software" section of QUIS (see Figure 2). Each item asked the participant to rank a different aspect of the interface on a 9-point semantic difference scale.

QUIS Overall Reactions to the Software Items		
Terrible	1 2 3 4 5 6 7 8 9	Wonderful
Difficult	1 2 3 4 5 6 7 8 9	Easy
Frustrating	1 2 3 4 5 6 7 8 9	Satisfying
Inadequate power	1 2 3 4 5 6 7 8 9	Adequate power
Dull	1 2 3 4 5 6 7 8 9	Stimulating
Rigid	1 2 3 4 5 6 7 8 9	Flexible

Figure 2. Items Included on Satisfaction Questionnaire.

Participants were also asked to indicate whether or not they had read the provided help text and whether or nor they would have clicked on a link to further help text had it been available. Finally, participants were given the opportunity to provide open-ended feedback about how they felt the interface could have been improved and to provide any other comments about the interface or the study (see Appendix B for the full questionnaire).

Efficiency was operationalized as "how long did it take for a user to create a metadata record for an image?" Time was recorded via system means. Web scripts logged the time a user first displayed the interface in his or her browser and the time a

user clicked the "submit" button to indicate the record has been completed. Efficiency was expected to be of limited use in assessing usability in the proposed study as users of the Definisions+Examples Version of the interface may require more time to complete their task simply because their interfaces provide more text for them to read. The efficiency measure is also more susceptible to any computer system-related problems that participants may experience than are the measures for satisfaction.

Testing Procedure

Participants were alternately assigned to two groups (A and B) based on the order in which they volunteered. Each group was asked to complete the steps below.

Step 1. Participants selected one of six images of botanical specimens

Step 2. Participants created metadata for the selected specimen, Group A using the Definitions Only Version of the interface, Group B using the Definitions+Examples Version of the interface

Step 3. Participants completed satisfaction questionnaire for the interface used in Step 2

Step 4. Participants selected a second image of a botanical specimen

Step 5. Participants created metadata for the selected specimen, Group A using the Definitions+Examples Version of the interface, Group B using the Definitions Only Version of the interface

Step 6. Participants completed satisfaction questionnaire for the interface used in Step 5

Testing took place remotely via the World Wide Web. Each participant was sent a URL for the study along with a study identification number to protect the integrity of the data and asked to complete the study at a convenient time.

Results

Study Participants

Twelve participants were recruited for the study. The 12 participants were evenly split in gender. Eight of the 12 were between 45 and 64 years old, with 2 between 18 and 24 and 2 over 65. All but one of the study participants responded that they use a computer and the Web daily. Nine of the 12 participants indicated that they have been using the Web for at least 4 years.

Each version of the interface was tested twice, resulting in 11 user tests for the Definitions Only Version (where on participant from Group B completed only the second of the two tests) and 12 user tests of the Definitions+Examples Version, for a total of 23 completed metadata records.

Satisfaction

Responses to the six QUIS satisfaction items were summed to provide a single measure of overall satisfaction with the interface with a minimum score of 6 and maximum score of 54[1]. The overall mean score for the Definitions Only Version was slightly higher than the Definitions+Examples Version at 32 compared to 31 (see Figure 3). The Definitions Only Version also scored slightly higher on the means for 5 of the 6 individual items; the Definitions+Examples Version scored higher only on the "Inadequate Power-Adequate Power" scale.

[1] In three cases participants did not provide a rating for the "Inadequate Power - Adequate Power" item. In computing the sum of the six items for those cases with missing data, the value for "Inadequate Power - Adequate Power" was imputed by using the mean score on that item for other participants in the same group when completing the same task.

	Definitions Only Version	Definitions+Examples Version
Overall Satisfaction	32.0	31.0
Terrible - Wonderful	5.6	5.4
Difficult - Easy	6.4	5.7
Frustrating - Satisfying	5	4.5
Inadequate Power - Adequate Power	4.7	5.5
Dull - Stimulating	5	5
Rigid - Flexible	5.3	5

Figure 3. Mean Scores on Satisfaction Scales by Interface

The mean difference between scores for the Definitions Only Version and the Definitions+Examples Version on any single item was never greater than 1, with only two items, "Difficult-Easy" and ""Inadequate Power-Adequate Power", having a mean difference of greater than .5. Also, neither the Definitions Only Version or the Definitions+Examples Version yielded satisfaction scores on any individual item greater than 6 or lower than 4, suggesting that participants didn't react extremely positively or extremely negatively to either interface.

A greater difference in satisfaction scores than that displayed between interface versions occurred between the participants' first and second tasks (i.e., the first and second time participants created a metadata record) regardless of the interface used. The mean overall satisfaction score for the second task was 4.9 points higher than that for the first task (see Figure 4). In addition, 4 of the 6 individual satisfaction items (all except "Rigid-Flexible" and "Dull-Stimulating") had scores at least .5 greater for the second task.

	1st Task	2nd Task
Overall Satisfaction	29.5	34.4
Terrible - Wonderful	5	6
Difficult - Easy	5.4	6.8
Frustrating - Satisfying	4.4	5.2
Inadequate Power - Adequate Power	4.7	5.7
Dull - Stimulating	4.9	5.2
Rigid - Flexible	5	5.2

Figure 4. Mean Scores on Satisfaction Scales by Task Order

As with the Definitions Only Version and the Definitions+Examples Version, testing order didn't elicit extremely positively or extremely negatively reactions.

While the sample size used for the study was too small to yield statistically significant results, the data suggest that the use of the metadata element examples in addition to metadata element definitions did not result in higher user satisfaction when compared to metadata definitions only. Instead, the data suggest a slight drop in user satisfaction.

Ten of 11 participants who used the Definitions Only Version said that they read the available help text and 10 also indicated that they would have clicked on a link to further help had it been available (see Figure 5). Ten of 12 participants who used the Definitions+Examples Version said they read the available help text, while all 12 said they would have clicked on a link to further help.

	Definitions Only Version	Definitions+Examples Version	1st Task	2nd Task
Did you read the available "Help" text?	10	10	10	10
If a link to further "Help" had been available, would you have clicked on it?	10	12	12	10
N	*11*	*12*	*12*	*11*

Figure 5. Number of participants answering "Yes"

Similar results were seen in comparing task order. The data clearly suggest that users desired more help than was available on either version of the interface tested.

Task Completion Time

The mean completion time for all task was 8:48. The shortest completion time was 1:15 and longest was 28:30. Because participants were not monitored while they completed their tasks completion times should be interpreted with caution. Participants on average took 4:15 longer to created a metadata record using the Definitions+Examples Version than they did using the Definitions Only Version (see Figure 6).

	Overall	Definitions Only Version	Definitions+Examples Version	1st Task	2nd Task
Mean Task Completion Time	8:48	6:35	10:50	11:02	6:22

Figure 6. Mean Task Completion Times

A slightly greater drop in mean task completion time was seen between the first and second task, with the second task completed on average 5:40 faster than the first task.

Participant Group A, which tested the Definitions Only Version then the Definitions+Examples Version, took on average slightly longer to complete their first task (8:21) than their second (7:53). On the other hand, Participant Group B, which tested the Definitions+Examples Version then the Definitions Only Version, saw a drop of over 7 minutes in mean completion time between their first and second tasks (See Figure 7).

	1stTask	**2nd Task**
Group A	8:21	7:53
Group B	12:57	5:06

Figure 7. Mean Task Completion Times by Group

The data are unclear as to whether or not metadata element examples increase or decrease task completion time. Group B, which was exposed to the examples in their first task, saw a much more dramatic drop in mean task completion time than Group B, which saw the examples during their second task. The data suggest the possibility that metadata element examples may initially increase completion time yet provide a decrease in completion time in subsequent tasks.

Qualitative Responses

To supplement that quantitative results reported above, a qualitative analysis was performed using the open-end responses to the questions "How could the previous form have been improved?" and "Other comments?"

The most prominent theme to emerge was that of "more help needed," occurring in the open-ended responses for 10 of the 23 total tests. For example, one participant said that he or she "[c]ouldn't find the 'Help' text" while another said simply that "it seems to

me that more guidance is needed on what to input in each field." Several responses requested specific types of guidance, including instructions on how to format names or a "checklist" describing a "prioritized listing of attributes."

Another theme to emerge from the open-ended responses was that of "more fields needed," occurring in the responses for 4 of the 23 responses. Several participants apparently felt constrained by the metadata fields made available to them. One participant wanted "[more] opportunities to describe the plant, such as habit, usage colors, etc." while another saw a need for "separate fields for scientific and common names." Other suggestions included leaf measurements and more detailed geographic information.

Many participants felt unclear about why they were completing metadata fields. These responses yielded a theme of "goal of system unclear," occurring with about the same frequency as the "more fields needed" theme. One participant said simply: "Didn't really get the purpose of the form as related to the illustration, which was complete enough in itself to make additional comments on the form (at least the info requested by the fo[r]m) seem useless." Other participants mentioned being "unclear relative to what is being solicited," another was "not sure I really understand the purpose of the form." One participant suggested that "it might be useful to provide some explanation of the ultimate goal of the effort -- what the final web-based product is intended to be."

Two participants in Group A suggested that examples of the metadata elements would improve the form following their use of the Definitions Only Version of the interface, with one participant after using the Definitions+Examples Version responding "[t]his was what I had in mind when I commented on the need for examples within the

help information." Two participants, however, responded negatively to the use of the examples in the help text. One said of the Definitions Only Version after having first used the Definitions+Examples Version "much better than the last." Another said "[this] was better than the first form, whose examples where less helpful than the short description of what was wanted" (this comment suggests that participants may have only skimmed the provide help text, rather than reading it in full).

Discussion and Future Research

Implications for Botanical Pride

Data collected in the study confirm Harmes's (2001) hypothesis that detailed help is an important factor in the usability of the tools for user-contributed metadata creation, including Botanical Pride. Only twice in the study did a participant indicate that he or she did not read the help text offered; only once did a participant answer that he or she would not have clicked on a link to further help had it been available. The open-ended comments reinforced the notion that participants desired more online help in using the interface. This desire for more help did not appear to be affected by the use of metadata element value examples.

In terms of both overall satisfaction and efficiency as measured in the study, the use of metadata element value examples seemed to provide no benefits over metadata element definitions alone (though it should be noted that the quality of the resulting metadata records was beyond the scope of this paper). While the usability of the Botanical Pride interface as measured by user satisfaction appeared to decline with the use of examples, the difference was not great. One interpretation of the data is that while

examples didn't increase user satisfaction, the didn't adversely affect it either. This, considered alongside participants' overwhelming preference for more help than was provided, would see to offer no reason to discontinue the use of examples in the Botanical Pride interface.

Though the Definitions+Examples Version of the interface provided longer mean task completion time whether it was tested on participant's first or second tasks, the shortest mean task completion times came on the second task of Group B, where participants used the Definitions Only Version after having first been exposed to the Definitions+Examples Version. Further, the difference in mean task completion times between groups using the Definitions Only Version and groups using the Definitions+Examples Version decreased from 4:36 on the first task to 2:47 on the second. The decrease in the difference in mean task completion time suggests that efficiency is affected more by the number of times a user has used the Botanical Bride interface than by the absence or presence of examples.

Despite the expressed desire for more help the mean scores on each of the satisfaction items was in the middle range of 4 to 6 on the 9-point scale. This suggests that online help may be a motivator, rather than a hygiene factor in Zhang and Dran's (2000) model. In other words, though more help was seem as beneficial, the available help was not seen as lacking enough to warrant extremely low satisfaction scores. This, along with the lack of substantial difference in satisfaction scores and the potential increase in efficiency time with use, again suggests that metadata element examples be included in some form in the online textual help of the Botanical Pride interface. While examples did not noticeably improve usability, the strongly expressed desire for more

help would seem to argue for testing different types and numbers of examples, rather than excluding them.

Implications for Further Study

Although the small sample size limits the generalizability of its results, the present study does offer some lessons for the design of future studies of the effect of definitions and examples in online help for tools for contributor-supplied metadata creation.

First, the comparatively large difference in mean task completion time between the first and second tasks suggests that future studies should include a higher number of tasks per user in order to better measure task completion times over time. For this reason, field tests, where participants use the tool repeatedly in their own work, rather than laboratory experiments may be more suitable for usability studies of metadata creation tools. The field test method has the additional benefit of placing metadata creation tasks within a work flow and context with which participants will be familiar (or grow familiar over time), alleviating the concerns expressed by participants in the present study about the "goals" or "purpose" of metadata creation. Moreover, a field test would offer more points at which to observe the affect of metadata examples and definitions, e.g., in initial metadata creation and during possible later metadata revision.

Second, future studies of the effect of metadata element examples and definitions on metadata creation tool usability should include a greater number of examples types (e.g., syntax, generic, and instance examples) and test interfaces. Increasing the number of interfaces tested would allow studies to gauge the effect of different types, numbers, and combinations of examples. For example, what is the optimum number of examples

to use in help for the Keyword or Subject element? Should generic and instance examples be used in combination or does one offer advantages over the other?

Finally, participant comments in this study included requests for more metadata elements, many of them specific to the Botany domain. This suggests that different knowledge domains may have quite different requirements for metadata creation tools, and begs the question of how a tool's available help could help address user concerns about the perceived inadequacy for the selected metadata scheme. Such a question may be especially important when the tool's target audience is composed of domain experts. Therefore, future usability studies of metadata creation tools may benefit from surveying tools from a variety of different knowledge domains.

Tools for contributor-supplied metadata are in many ways still in their infancy with respect to the Web. Clearly, greater demands will be placed on such tools as more and different communities recognize the need to utilize resource authors and contributors as metadata creators. This study demonstrates that textual help is a crucial factor in the usability of such tools and that further research examining metadata element examples is needed to determine how to best provide it.

References

Beyer, H & Holtzblatt, K (1998). *Contextual design: defining customer-centered systems*. San Francisco: Morgan Kaufman.

Burnett, K, Ng, K, & Park, S (1999). A comparison of the two traditions of metadata development. *Journal of the American Society for Information Science*. 50, 1209-1217.

Chin, J, Diehl, V, & Norman, K (1988). Development of an Instrument Measuring User Satisfaction of the Human Computer Interface. *ACM conference proceedings on Human factors in computing systems.*

Dublin Core Metadata Initiative (1999). Dublin core metadata element set, version 1.1: reference description. Retrieved November 1, 2001 from http://dublincore.org/documents/dces/

Duval, E, Hodgins, W, Sutton, S, & Weibel, S (2002). Metadata Principles and Practicalities. D-Lib Magazine, 8:4. http://www.dlib.org/dlib/april02/weibel/04weibel.html

European Usability Support Centres (2001). Usability Standards. Retrieved October 14, 2001, from http://www.lboro.ac.uk/research/husat/eusc/r_usability_standards.html

Fraser, B & Gluck, M (1999). Usability of Geospatial Metadata or Space-Time Matters. *Bulletin of the American Society for Information Science*, 25:6. Retrieved November 20, 2001 from http://www.asis.org/Bulletin/Aug-99/fraser_gluck.html

Frøkjær, E, Hertzum, M, & Hornbæk, K (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the CHI 2000 conference on Human factors in computing systems. April 1 - 6, 2000, The Hague Netherlands*

Greenberg, J, Pattuelli, M, Parsia, B, & Robertson, W (2001). Author-generated Dublin Core Metadata for Web Resources: a baseline study in an organization. *Journal of Digital Information*, 2:2. http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/

Hanrath, R (2002). An analysis of online help in interfaces for contributor-supplied metadata. Unpublished paper.

Harmes, H (2001). Development of an Input Form to Capture Author-Generated Metadata for a Botanical Image Collection. Unpublished Master's thesis.

Hert, C & Jacob, E (2000). A usability study of online indexing structures in the networked environment. *Journal of the American Society for Information Science*, 51, 971-988.

Kumar, V, Furuta, R, & Allen, R (1998). Metadata visualization for digital libraries: interactive timeline editing and review. *Proceedings of the third annual ACM conference on digital libraries.*

Markey, K (1984). Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research*, 6, 155-77.

Miller, E (1998). An introduction to the Resource Description Framework. *Bulletin of the American Society for Information Science*, 25:1. Retrieved October 15, 2001 from http://www.asis.org/Bulletin/Oct-98/ericmill.html

Milstead, J &Feldman, S (1999). Metadata: cataloging by any other name. *ONLINE*, January 1999. http://www.onlinemag.net/OL1999/milstead1.html

Nielsen, J (1994). Heuristic Evaluation. In J. Nielsen and R. Mack (Eds.), *Usability Inspection Methods* (pp 25-62). New York: John Wiley & Sons.

Nielsen, J (2000). *Designing web usability*. Indianapolis: New Riders.

J Nielsen & R Mack (Eds.). (1994) *Usability Inspection Methods* New York: John Wiley & Sons.

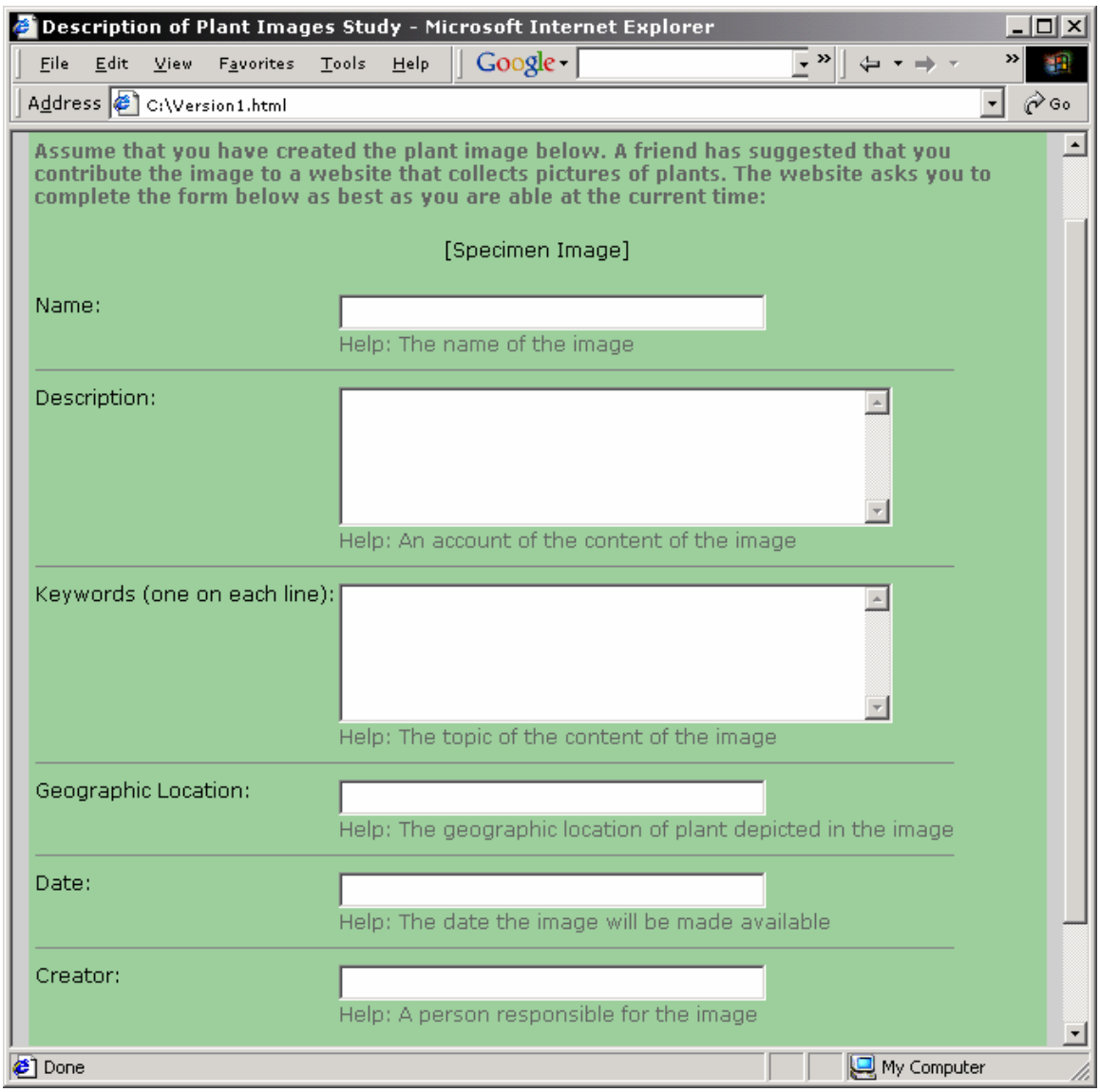
Rubin, J (1994). *Handbook of usability testing: how to plan, design, and conduct effective tests*. New York: John Wiley & Sons.

Thomas, C. & Griffin, L. (1999) "Who will create the Metadata for the Internet?" *First Monday*, 3:12. http://www.firstmonday.dk/issues/issue3_12/thomas/index.html

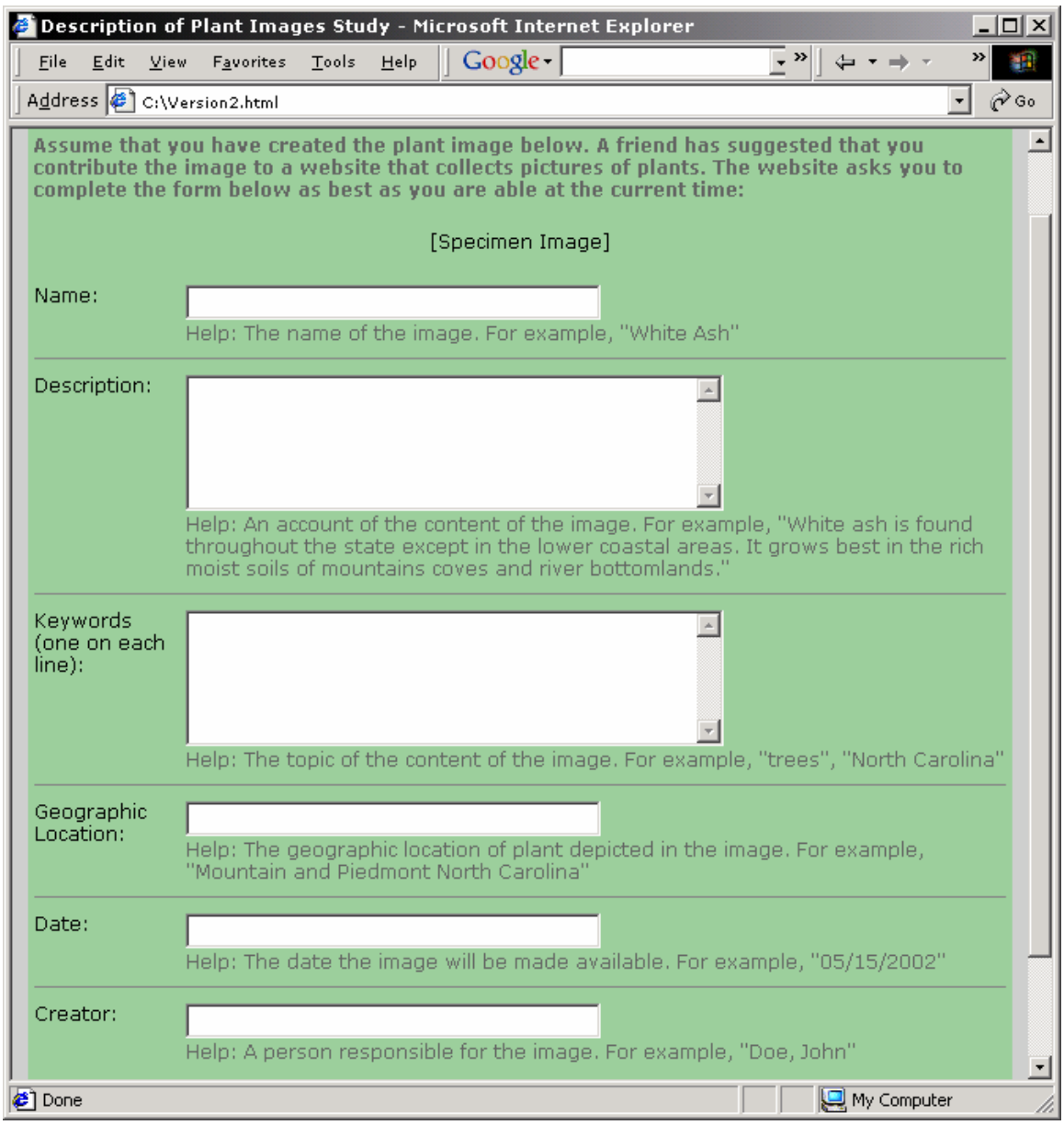
Zhang, P & von Dran, G (2000). Satisfiers and dissatisfiers: a two-factor model for website design and evaluation. *Journal of the American Society for Information Science*, 51, 1253-1268.

Appendix A. Interfaces Tested

The Definitions Only Version



The Definitions+Examples Version



Appendix B. Study Questionnaire

1. Please rate the previous form using the following criteria:

Terrible	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Wonderful
Difficult	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Easy
Frustrating	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Satisfying
Inadequate power	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Adequate power
Dull	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Stimulating
Rigid	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	Flexible

2. Did you read the available "Help" text?

[ ] No
[ ] Yes

3. If a link to further "Help" had been available, would you have clicked on it?

[ ] No
[ ] Yes

4. How could the previous form have been improved?

5. Other comments