Partnership Reviewing: A Cooperative Approach for Peer Review of Complex Educational Resources

John Weatherley[1], Tamara Sumner[2], Michael Khoo[3], Michael Wright[4], Marcel Hoffmann[5]

[1] DLESE Program Center
[4] Unidata Program Center
UCAR, P.O. Box 3000
Boulder, CO 80307 USA
[1] +1 303 497 2680
jweather@ucar.edu[1]
mwright@ucar.edu[4]

[2] Dept. Computer Science
[3] Dept. Communication
University of Colorado
Boulder, CO 80309 USA
[2, 3] +1 303 492 1592
sumner@colorado.edu[2]
khoo@colorado.edu[3]

[5] Informatics and Society
University of Dortmund
Germany
+49 231 755 4715
marcel.hoffmann@udo.edu[5]

ABSTRACT
Review of digital educational resources, such as course modules, simulations, and data analysis tools, can differ from review of scholarly articles, in the heterogeneity and complexity of the resources themselves. The Partnership Review Model, as demonstrated in two cases, appears to promote cooperative interactions between distributed resource reviewers, enabling reviewers to effectively divide up the task of reviewing complex resources with little explicit coordination. The shared structural outline of the resource made visible in the review environment enables participants to monitor other reviewers' actions and to thus target their efforts accordingly. This reviewing approach may be effective in educational digital libraries that depend on community volunteers for most of their reviewing.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*computer-supported cooperative work, collaborative computing, asynchronous interaction;* H.3.7 [**Information Storage and Retrieval**]: Digital Libraries—*collection, standards.*

General Terms
Human Factors, Design, Management, Experimentation.

Keywords
Computer-Mediated Communication, Distributed Cognition, Educational Digital Libraries, Peer Review, Scholarly Publishing, Second Order Interactional Effects

1. INTRODUCTION
Distinguishing features of a digital library are the policies it puts in place to manage its collections [3]. These include accessioning policies (governing what gets chosen for inclusion), de-accessioning policies (governing what resources are removed from the library), and management and maintenance policies crafted to maintain the evolving conceptual and physical integrity of collections. Peer reviewing policies are emerging as an important tool for managing the conceptual integrity of digital collections. Previously, with traditional 'bricks-and mortar' libraries, peer reviewing occurred outside the confines of the library within academic communities defined by journal circulations and conference participants. Now, many digital libraries see peer review as an important service that adds value to both their collections and their user communities. The perceived values include providing a form of quality control and promoting a shared activity for community participation in collections management [23].

We are investigating a particular form of technology-supported peer review: the Partnership Review Model. This model has been crafted both technically and socially to meet these dual quality control and community participation goals. Using a specially designed anchored discussion forum, groups of reviewers and authors work together to critique and discuss digital resources. A key feature of this review environment is a Publisher's Toolkit, which parses a web resource to create a structural outline of the digital resource to be reviewed [34]. This structural outline provides the anchors for a threaded discussion environment and serves as a visible, shared representation of the evolving review process for all participants.

In this article, we present the Partnership Review Model and its underlying technology support in the context of two cases: peer review of hypermedia course modules (JESSE) and data visualization software (UMADA). JESSE—The Journal of Earth System Science Education—is one of three peer review processes being developed in the Digital Library for Earth System Education (DLESE) [24]. The other two processes are described elsewhere [19-21]. UMADA—Unidata Meteorological Applications Discussion Area—is using the Partnership Review Model to enable distributed user participation in the design of next generation scientific data analysis and visualization software tools. Some of the visualization clients created by UMADA will be integrated into digital libraries such as DLESE and the National Science, Technology, Engineering, and Mathematics Digital Library (NSDL) to enable learners to access library resources containing earth system data. In both cases, the project initiators envisioned using the Partnership Review Model to support collaborative dialogs between participants, with the ultimate aim of generating knowledge useful for improving resource design.

JCDL'02, July 13-17, 2002, Portland, Oregon, USA.

The research presented here builds on our prior work on fostering collaborative peer review of scholarly articles in multidisciplinary electronic journals [5, 34, 35]. This work investigated how new forms of technology-supported peer reviewing could be used to promote collaborative, dialogic exchanges between reviewers and authors. Analyses of these dialogs revealed that authors and reviewers were using the peer review system to negotiate the meanings of underlying concepts, to discuss what constitutes appropriate research methodology and how best to present research results. We postulated that these types of collaborative dialogs were important for both community building in distributed environments and for creating shared understandings in multidisciplinary fields such as educational technology.

Here, we extend this line of inquiry to examine peer reviewing of complex digital resources, such as large hypermedia course modules and scientific data visualization software. Reviewing these types of objects has both similarities and differences to reviewing scholarly articles. One key difference arises from the complexity and heterogeneity of the objects to be reviewed. While scholarly articles discuss complex ideas, the articles themselves are of relatively short length and predictable structure: a reviewer can usually read a scholarly article in less than a day and does not have to expend time or cognitive effort to learn how to 'use' or navigate the article. The resources considered here are both unpredictable in structure and of high functionality; i.e., their sheer size, number of media elements, or number of functions makes it impossible for any single person working independently to review the entire resource in a reasonable amount of time. For some very complex software artifacts, functional coverage is considered to be impossible [12].

An interesting finding from the cases presented here is that the affordances of the review system, namely the explicit, shared structural representation, appeared to support *emergent cooperative behavior*. Without explicit coordination or negotiation, participants effectively divided up the task of reviewing large, complex objects resulting in reasonable coverage of the resources' contents and functionalities.

In the remainder of this article, we begin with an overview of recent innovations in peer reviewing approaches and discuss some of the unique needs of educational libraries and educational resources for reviewing systems. The core of this article presents the Partnership Review Model and the two cases. We conclude by critically examining our empirical evaluations of, and experiences with, using the Partnership Model in these cases. We hope that this article contributes to the knowledge in our field by (1) providing a concrete presentation and evaluation of an innovative cooperative approach to peer review, and (2) identifying and discussing where and how the needs for educational object review may differ from scholarly article reviewing.

2. PEER REVIEW

Histories of peer review place its emergence in Britain and France in 1665, in the *Journal des Sçavans* (Paris) and the *Philosophical Transactions of the Royal Society* (London). In time, these publications came to be considered by scientists as places to establish intellectual precedence and the claim of the author as discoverer. The journals promoted their own authority (and thus also the flow of submissions) by developing quality assurance and systems of vetting that became institutionalized in reviews by 'peer' members of the scientific community. In the present form of peer review, authors submit manuscripts to journals, the manuscripts are sent out by the editor(s) to reviewers, the reviewers evaluate the manuscript and return it to the editor, and then the editor makes the decision whether or not to publish, often passing the reviewer's comments back to the author [13, 25, 37].

This process assumes peer reviewers to be members of a scientific community with shared perceptions, each of whom are capable of rationally evaluating the worthiness of a submission from the same point of view, and whose task is to make evaluative judgments. A corollary of this assumption is that the objects of peer review are themselves capable of being reviewed by the members of this scientific community [32]. However, several studies of inter-reviewer reliability indicate that the scientific communities presumed to underpin journal publication and peer reviewing can be more heterogeneous than is often assumed, especially in the social sciences, which experience higher levels of rejection than the natural sciences. From a Kuhnian point of view, this can be explained as a consequence of the lack of cohesive dominant paradigms in many social scientific fields, within which social scientists can agree how to 'do' science [22]. Inter-reviewer reliability (agreement between reviewers) can also be affected by factors which reviewers can respond to in individual ways, such as the author's gender, age, rank, affiliated institution, and practices such as self-citation [7-9, 14, 33]. Peters and Ceci [27] resubmitted under different names twelve papers previously accepted by influential American psychology journals; although the resubmissions were to the same journals that had previously accepted them, eight were now rejected for serious 'methodological flaws.' Finally, inter-reviewer reliability can correlate with the number of reviewers used, with reliability decreasing as more reviewers take part in the process. To counter these and other biases, researchers have suggested such practices as wider use of blind reviews, and conversely, signed reviewing to hold reviewers accountable for their comments [7, 15, 27]. In the context of scientific community heterogeneity, however, these suggestions attempt to overcome rather than engage with this diversity.

Peer review reliability is also problematic in electronic publications. Studies indicate that electronic publications lack credibility among academics, at least partly because they are perceived to be less rigorous in their academic quality control than their paper counterparts [2, 31, 33]. Electronic journals have attempted a number of strategies to reduce this mistrust, ranging from reproducing paper review processes in computer-mediated communication (CMC) forms, to exploring new possibilities afforded by CMC [15, 16]. These latter attempts include e-mail supported 'round robins' [28] collaborative dialogs using threaded discussion forums [5], and open review and ratings systems coupled with collaborative filtering [30]. Harnad has advocated strategies that involve facilitating ongoing discussions between authors, reviewers, and editors, to produce 'scholarly skywriting.' The mushrooming diversity of these strategies indicates that while peer review is considered necessary for electronic publications, its form is not necessarily fixed. Siemens et al.'s survey of 336 Canadian academics found that while a majority of respondents (81%) wanted a form of guarantee equivalent to that of paper publications, a majority (62%) also said that they were interested in exploring CMC based solutions to achieve this [31].

The peer review concerns of interdisciplinary social science and of electronic publication come together when considering how to review the educational resources now being cataloged in digital libraries such as DLESE. On the one hand, the need for assured quality of DLESE resources as a precondition for user satisfaction is repeatedly emphasized in DLESE discourse [19, 20, 23]. On the other hand, DLESE's reviewing community is very diverse (comprised of, for example, teachers, scientists, pedagogists,

librarians, and quality assurance testers). Plus, the educational objects considered by DLESE are heterogeneous, complex, and often large scale entities that include hypermedia course modules, models and simulations, software tools, virtual field trips, teacher's guides, assessment modules, and so on. Outside the digital arena, a review of complex educational resources, namely science textbooks, recently completed by AAAS took a large (paid) multidisciplinary team two years to complete [1]. The large effort and variety of expertise required to review complex educational objects is of particular concern for digital libraries that depend on the volunteer efforts of users to conduct resource reviews.

To deal with community and resource diversity, DLESE is pursuing a number of alternate and complementary strategies, referred to as 'alternative pathways to the reviewed collection' [23]. In the next section, we describe work related to one of these pathways, the Partnership Review Model. Analyses of reviewers' activities in two cases (JESSE and UMADA) suggest that the review forum supported emergent second order interactional characteristics among reviewers. Participants used the forum not just for the original (first order) design intention of communicating and collaborating with other participants, but also to monitor the task progress of other participants. These second order characteristics, it will be argued, leverage what is conventionally seen as a disadvantage in peer review—the multidisciplinary institutional expertise of participants—against complex multimedia educational objects. In describing the second order effects observed in JESSE and UMADA, a theoretical and practical distinction will be made between 'cooperative' and 'collaborative' work.

3. THE PARTNERSHIP REVIEW MODEL

The Partnership Review Model [36] has been used to implement and analyze technology-supported peer review processes in which reviewers and authors work together to collectively critique and evaluate a digital resource. The model's ideals and goals are informed by several related theoretical approaches. Hutchins' theory of distributed cognition [18] is used to inform the model's analytic framework. Putnam's notion of social capital and Axelrod's definition of cooperative behavior inform the model's implementation in specific settings.

Putnam describes social capital as "features of social organizations such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit" [29]. As such, social capital embodies the belief that people working *collaboratively* toward a common goal can achieve more through combining their expertise and knowledge than the same individuals working alone. Axelrod identifies three conditions as being necessary for *cooperative* behavior to thrive in groups [4]. First, individuals must be able to identify each other within the group. Second, members must have information on others' behavior. Third, it must be likely that individuals in the group will meet again, otherwise there is little motivation for cooperation.

It is important to note that collaborative behaviors and cooperative behaviors are not the same thing. According to Dillenbourg et al., "cooperation and collaboration do not differ in terms of whether or not the task is distributed, but by virtue of the way in which it is divided." Cooperative tasks can be split into independent subtasks, with coordination mainly required when assembling the partial results. In collaboration, processes for carrying out the task are intertwined and coordination is ongoing as participants continually attempt to construct and maintain a shared conception of a problem [11].

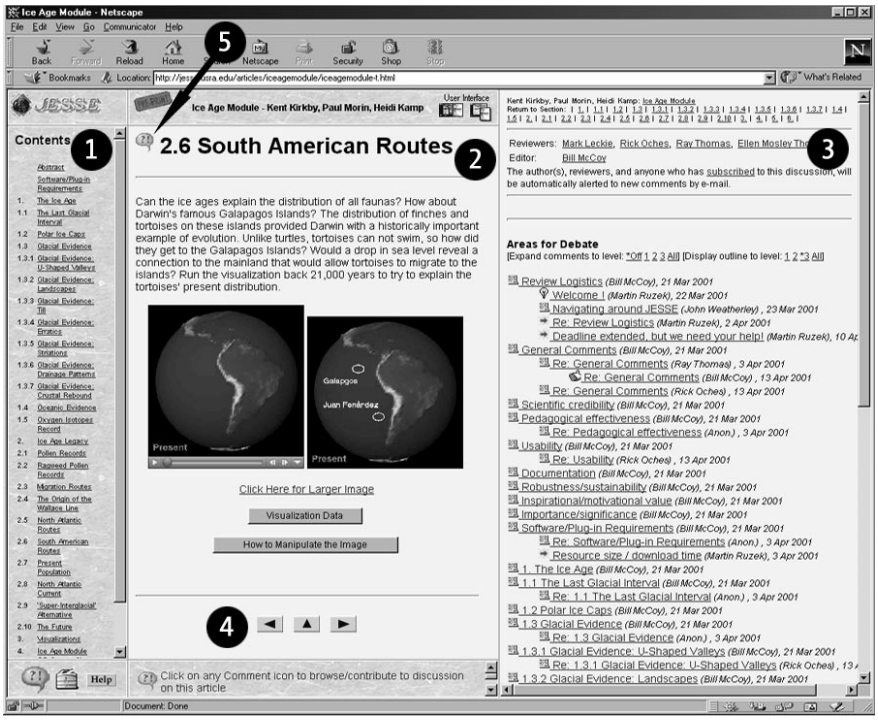

Figure 1. The Partnership Review Environment. (1) Table of contents. (2) Body frame holds the artifact. (3) Review comments frame where general review topics, artifact's structural outline, and review activity are visible. (4) Forward, back, top buttons. (5) Discussion anchor opens the review comments for this section in frame 3.

The Partnership Review Model describes a social and technical environment that encourages both *collaboration* and *cooperation* among reviewers and resource authors/designers, through the use of a specially designed HTML-based anchored discussion forum known as the Digital Document Discourse Environment (D3E). In this environment, the resource itself, or a representation of the resource, is provided with embedded anchors that link specific sections in the resource to related review discussion. D3E includes a Publisher's Toolkit that parses the resource to identify sections or key structural elements to create the anchored environment shown in Figure 1. D3E is one of the earliest examples of anchored discussion environments that provide a tight coupling between the artifact and its related discourse. Other examples include Churchill et al.'s Anchored Conversations, which allows users to anchor chat dialog boxes to documents in Microsoft Windows applications like Word, PowerPoint and Photoshop [6].

Each comment within the D3E environment includes the participant's name and email, thus supporting Axelrod's first condition. When a comment is submitted, all participants are notified automatically by

email, indicated the level and frequency of review activity as it is in progress. The environment places the review discussion next to the resource itself in an adjoining web browser frame, creating a close coupling between the resource and its review. The visible, shared structural outline created by the environment, and the email notifications to a lesser degree, serve to meet Axelrod's second condition by enabling participants to easily monitor other participants' actions and the evolving review process as a whole. While not demonstrated in the two cases presented here, as multiple reviews are accumulated in the collection, searching and browsing tools may be used to provide a visible history of reviewer participation. Such additional tools can help gather, sort, and reveal each individual's contributions over time, thus indirectly supporting Axelrod's third condition.

The social process component of the model is best explained by concrete example. We present two cases where the Partnership Review Model was used in real settings for extended periods of time. For each case, we begin by outlining the overall context of the cases: what were the goals, who were the participants, and what types of objects were being reviewed. We then describe the review processes, both as originally envisioned by the project initiators, and what actually occurred as revealed through our data analyses.

4. RESEARCH METHODOLOGY

Distributed cognition [18] theory underpins the research methodology used to study the Partnership Reviewing processes in our cases. The theory of distributed cognition provides a framework for analyzing interactions in cooperative work as the propagation of representational state across individual participants, representational media, and time [17]. Distributed cognition aims at developing an understanding of social practice in distributed environments by focusing on activities' relations to artifacts and on the analysis of their social organization. Hutchins combines different methods to study distributed cognitive systems, namely artifacts or "material analysis," analysis of the spatial layout of work settings, and more extensively the ethnographic approach of engaging in long-term participant observation [18]. Our methodology resembles Hutchins' recommendations to a great extent.

Participant Observation. Two of the authors (Weatherley and Wright) were participants in the projects. In JESSE, Weatherley worked with the founding editors to design, deploy, and support the initial field trials. These trials lasted approximately six months. In UMADA, Wright participated in the project for 18 months, assisting to set up and facilitate the distributed process from shortly after the project's beginning. Since some UMADA participants are located in Boulder, CO, we were also able to gather data from observing participants using the discussion forum or performing other activities. As with other research studying asynchronous online collaboration, many observations in both cases were limited to looking at the discussions taking place in the computer-mediated environment. These included discussions currently taking place and those captured in the UMADA project's extensive archives.

Artifact Analysis. We carried out a broad quantitative analysis of the review discussions in order to characterize the overall amount and pattern of participation. We also conducted detailed content analyses of selected discussions to better understand the emergent interactional characteristics of the review forums. In JESSE, we analyzed all of the review discussions, namely the original field trial and the subsequent review. In UMADA, we chose to analyze four of the archived discussions that participants regarded as either least or most successful in terms of active participation and review outcomes. In the detailed analysis, we examined the extent of collaborative dialog among project participants, identifying who participated in specific discussion threads, which contributions were monologic statements, and which were contributions to a dialog between multiple participants. Next, we analyzed these selected discussions to determine the outcomes of the review processes. In both cases, the participants were ostensibly engaging in the high-level cognitive task of making evaluative judgements; they were supposed to note problems or deficiencies in the resources being reviewed and to make suggestions for improvements. To study these outcomes, we noted the overall number of judgements made, who made them, and how many were unique or had been proposed elsewhere. Finally, given the complexity of the resources being reviewed, we also examined the coverage of the review discussions. We defined coverage in terms of the structural outline of the resource made visible in the review environment. We noted the percentage of elements in this outline where evaluative judgements had been made for individual participants and for the review groups as a whole.

Semi-structured Interviews. We also interviewed many of the participants in both cases. In JESSE, five participants were interviewed over the phone. The interviews lasted from 45 to 60 minutes each and detailed notes were taken. In UMADA, six participants were interviewed face-to-face, enabling us to jointly examine the discussion archives during the interviews. These interviews lasted from 60 to 90 minutes each and were taped and transcribed.

5. CASE 1: JESSE

JESSE is an emerging multi-disciplinary journal for geoscience and related disciplines involved in Earth system and global change education (http://jesse.usra.edu/). The goal of the founding editors of JESSE is to provide a reviewed collection of diverse earth system science educational resources. JESSE is one avenue being considered for bringing such reviewed resources into DLESE. The scope of these resources includes course modules, lesson plans, models and simulations, pedagogical and resource design papers, and so on, with most being available exclusively in electronic forms. Because of its ability to accommodate diverse electronic resources as well as textual material, the editors chose to use D3E as the reviewing tool.

The resource chosen for the first review, the Ice Age Module, is a hypermedia course module targeted at high school and undergraduate introductory geoscience education. It details the history of glaciers and climate change. This media rich module is 138 Megabytes in size, and consists of approximately 50 HTML pages, 12 images, 39 interactive visualizations, and 16 interactive question and answer quizzes. The module as it was displayed in the review environment is shown in Figure 1. The review of this module took place over a three-week period and involved eight participants (six reviewers, and two resource authors). The resource chosen for the second review was an article on using computationally intensive models in the classroom; embedded links were provided to enable reviewers to familiarize themselves with some of the complex models discussed prior to the review. The second review trial took place in August 2001, and involved 4 participants (3 reviewers, and 1 author). A broad characterization of participation in these reviews is shown in Table 1.

Research in technology frames suggests that individuals show a tendency to approach new tasks using assumptions gained from prior experience [26]. Because of this, three e-mail

Table 1. Overall characterization of participation in JESSE.

	Review One	Review Two
Participation:		
Total number who made comments	**4**	**3**
Reviewers	**4**	**2**
Authors	**0**	**1**
Total number of comments	**24**	**31**
By reviewers	**24**	**16**
By authors	**0**	**15**
Number of monologic comments[1]	**17**	**1**
Number of dialogic comments[2]	**7**	**30**
Number of two person dialogs	**2**	**15**
Number of three person dialogs	**1**	**1**
Coverage:		
Total number of anchors	**37**	**19**
Number of anchors with comments	**19**	**15**
Percent of anchors with comments	**51%**	**79%**

[1] Individual comments not involved in any dialog.

[2] Individual comments involved in dialog between two or more participants.

communications were made to participants at various stages in the review process to help distinguish the Partnership Review approach from traditional review processes with which they were familiar. The first email was sent about a week before the review began and was essentially a 'welcome' letter introducing the participants to each other and describing the review process at a high level. The second 'process-oriented' email was sent shortly before the review commenced; participants were encouraged to work collaboratively rather than individually, and were asked to begin commentary early in the review period, providing opportunity for continuing dialog to take place over the remainder of the period. The third email was sent shortly before the review period ended to remind participants to wrap up. The motivation for these e-mails was to explicitly encourage participants to begin to reframe their actions and expectations in light of the new work practices inherent in the Partnership Review Model. These forms of explicit intervention or mediation have been shown to be effective at supporting changes in work practices in other reviewing contexts [35].

Collaboration. The Ice Age Module trial contained 24 review comments, seven resulted in three dialogic exchanges; the other comments were not replied to (monologic). Overall there was very little dialog in the first review. Twelve of these comments offered positive critiques on aspects of the resource; 2 offered negative critiques, and the remaining were neutral statements offering factual information or rationale. In the second review, there were 31 comments made by two active participants, resulting in 15 short two-comment reviewer-author dialogic exchanges. Thus, in neither review did we observe the collaborative, knowledge-generating exchanges originally envisioned by the project initiators (founding editors). This suggests that one original goal of the Partnership Review Model, to create a collaborative review exchange, generally was not met in these cases. In the interviews, participants felt this review approach to be more informal than traditional models, but still viewed it as offering comparable quality assurance. Some participants noted that they were "more polite" in how they worded their comments; another participant expressed the opposite perspective. He felt he could be more "picky" since authors would have an opportunity to clarify their positions.

Cooperation. An analysis of the pattern of reviewer activity revealed a tendency for reviewers to focus on different parts of the resource, as made visible in the review environment. In the first review, while 19 of the 37 anchors (51%) had review comments, 16 of these contained comments from only one reviewer. One reviewer made 14 comments covering approximately 38% of the review topics while the other three participants made fewer comments (between one and five each). However, each reviewer generally focused on different areas, resulting in complementary coverage of a little over half of the complex resource. Interviews revealed that reviewers were indeed aware of others' commentary and the areas where review was taking place. Several reviewers expressed concern with the time commitment required to review this complex resource. One reviewer noted however that the Partnership Review Model provided an opportunity for individuals to bring their particular expertise to the table, relieving any one reviewer of the responsibility of having to make a comprehensive review.

Coverage. In the second review, only one reviewer and one author participated, hence we observed little reviewer cooperation here. However, this reviewer chose a review strategy that yielded nearly full coverage of all parts of the resource as outlined in the review environment. It may be that, seeing the breadth of his work, other reviewers felt their contributions weren't necessary. Further interviews could shed light on these behaviors. It does seem as if the visible structural resource outline encourages reviewers to spread their activities broadly across complex resources, resulting in reasonable coverage (51% and 79%) of large resources.

6. CASE 2: UMADA

The Unidata program provides software tools, technical support, and data distribution services to atmospheric science researchers and educators. In 1998, Unidata initiated a project to develop the next generation of platform-independent meteorology data analysis and visualization applications, called 'MetApps', to replace the current FORTRAN and C-based systems. Towards this end, a MetApps task force was established consisting of 14 user-volunteers from the community and five Unidata staff developers. The developers are all located at the Unidata Program Center in Boulder, Colorado; the users are geographically distributed throughout the United States. Most users are university faculty engaged in teaching or research in atmospheric science; a few are from government-run meteorological centers.

An explicit goal of the project was to adopt a process that: (1) enabled users to meaningfully participate in the design process, including requirements definition, and (2) helped to promote a sense of community responsibility for the project outcomes. Towards this end, the group developed a communication strategy to enable geographically distributed users to participate in the design process. They decided to rely on 'use case' descriptions and functional prototypes as the major design representations during the project. Use cases use natural language to describe how different types of users might interact with the system and are a popular approach to requirements definition among object-oriented software developers [10]. In MetApps, all use case descriptions are developed and written by users. Staff developers are responsible for implementing functional prototypes based on all or part of use cases selected by the group for implementation.

The UMADA site consists of a front page, and a growing number of current and past discussion forums. The D3E Publisher's

Table 2. Distribution of discussions across time and participation across discussions. TF is total number of contributions from task force members. D refers to contributions from staff developers.

	03/99						12/99			01/00									12/00			ΣTF	ΣD
Discussion of Applications																							
Interavtive Sounding (V1)																						1	0
Interactive Sounding (V2)																						85	22
Interactive Sounding (V3)																						18	11
Interactive Sounding (V4)																						37	16
Surface Observation																						35	12
Gridded Data Viewer (V1)																						21	19
GDV (V2,3 & 4)																						29	24
Image Viewer																						14	11
Discussion of Use Cases																							
Model Data Viewer																						4	0
Satellite Viewer																						15	2
Radar Data Viewer																						0	0
Meta-discussions																							
Documentation & Support																						0	0
Future Plans																						6	1
UMADA vs. email																						4	1
Conference Call																						5	8

Toolkit is used to create anchored review forums for both types of resources (uses cases and functional prototypes), where threads are automatically created that correspond to specific parts of the resource being discussed. The web-based interface created by the toolkit looks similar to that shown in Figure 1. During a period of about two years, developers and task force members used UMADA to discuss three use cases, nine releases of four different prototypes and to conduct four meta-discussions about issues such as future project directions and UMADA usage. Table 2 shows the general pattern of participation during this period.

During the use case stage, users develop use case descriptions and publish those in the UMADA environment for discussion by the developer team and members of the task force. Usually the use case stage is followed by a prototype stage. Developers are responsible for prototypes and their discussions. The prototype's associated documentation is used to generate the review environment's structural outline, which has the benefit of forcing the documentation to be developed and reviewed in conjunction with the prototype. Users are responsible for installing, experimenting with, and reviewing the prototype; they report that the software is usually complex, taking at least two weeks to install and explore.

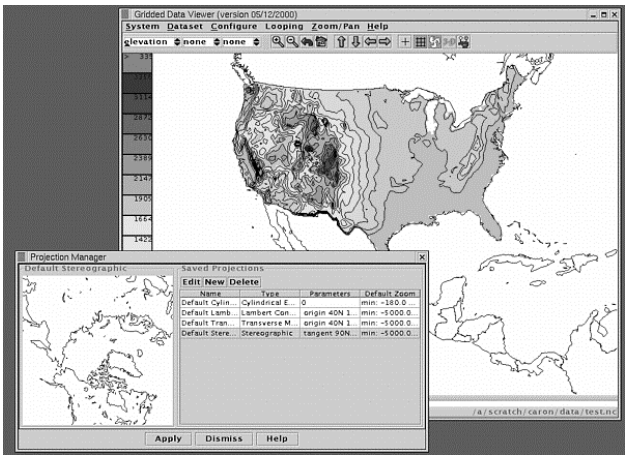

Figure 2. Gridded Data Viewer Prototype.

We will use the example of the Gridded Data Viewer (GDV) to illustrate some of the review behaviors observed. The GDV prototype displays and manipulates data on grid point systems (e.g., latitude and longitude), and is a representative application created by the project (Figure 2). It arose from four Model Data Viewer use cases contributed by different task force members. This discussion had 53 contributions. The GDV discussion is especially interesting since it shows a group of experts discussing a resource for the second time. The group is—at this point—already experienced in using the UMADA environment. Six task force members and three staff developers participate. We have a good distribution of contributions among the participants, with no one dominating the discussion.

Collaboration. When analyzing the GDV discussions, two interesting patterns emerged that appear to be repeated in most of the UMADA discussions. Of the 53 comments, 45 were dialogic in nature and the remaining were monologic statements. We identified 18 separate dialogs consisting of four two-way, four three-way, and two four-way conversations. Figure 3 shows an excerpt from a three-way discussion where users and developers collaboratively constructed a new design requirement. There were no 'whole group' discussions involving all nine participants. This precluded the establishment of a larger consensus for the group as

Re: 2.0 Features *(TaskForceLeader, tfl@mail.edu) 22 Oct 2000 16:57 MDT*

Since the 3D component of GDV is giving ross sections in two directions (detemined by grid) can the 2D gdv do the same instead of just the quasi-north south cross sections. For so called quick and dirty answers one may not wish to incur the overhead of 3D but would still like cross sections in two perpendicular directions.

Re: 2.0 Features *(Developer2, d2@ucar.edu) 23 Oct 2000 11:00 MDT*

Do you mean an arbitrarily oriented vertical cross section, rather than the current cross section that is always a y-z data slice (x = constant) ?

If so, it it still a single plane, or does it need to be segmented (eg for an airplane flight path?).

Any ideas on how the user would specify what they want?

Re: 2.0 Features *(TaskForceMember3, tfm3@mail.edu) 24 Oct 2000 09:45 MDT*

I'd say first lets build in the ability to just draw an arbitrarily oriented straight cross section. Perhaps there can be a cross section tool. The user selects the tool, then clicks and drags to define the two endpoints of the cross section. Once the cross section is defined, the user should be able to drag the endpoints if they want to change the location of the cross section.

A vertical scale on the cross section would be nice. So would the abilility to show only a portion of the domain (in both x, y, and z) David

Re: 2.0 Features *(Developer2, d2@ucar.edu) 24 Oct 2000 17:13 MDT*

these are good ideas that could quickly improve the xsection feature.

i'll think about how to do them.

Figure 3. Reviewers and developers discuss cross section feature

a whole. We found that developer participation is the single most important factor determining whether a posted comment sits unreplied to or turns into a dialog. It appears that users rarely engage in discussions only with each other. At this point, we can only speculate as to why this is so. One possibility is that it is symptomatic of a perceived lack of ownership in the product; i.e., it is perhaps only motivating to discuss changes to the resource with those empowered to actually implement your propositions. A consequence of this pattern is that, for dialogs to occur, the onus is on developers to stay on top of the discussion and respond rapidly to user comments. This level of participation is very time consuming and requires considerable commitment from developers.

Cooperation. Task force members made 49 comments (requirements, bug reports, etc.) related to the prototype; developers responded directly to 35 of them by either discussing the issue, or by promising to put the suggestions into the next or future release. Of these 49 comments, we identified 23 newly proposed design requirements, of which three were directly incorporated into the next release and eight more were incorporated into future development plans. Five were discussed further and the remaining were not followed up in the discussion forum. Our analyses revealed two interesting aspects to these requirements. First, of the 23 total requirements identified, 21 were unique. Two were simply paraphrases of the same requirement; and two articulated opposite positions on the same issue. This level of 'uniqueness' seems extremely high to us, suggesting that participants are tracking the suggestions and only contributing new ones. The UMADA forum appears to act as a visible 'group memory' that helps to coordinate distributed cooperative actions; i.e., without explicit negotiation or coordination, reviewers appeared to efficiently divide up the areas of the prototype to focus their efforts on. As one interviewee put it, the review environment enabled everyone to focus on the issue they personally were most interested in.

Coverage. This use of the environment to support cooperative behaviors was further underlined in our coverage analysis of developers' comments. While three developers participated in the overall discussion, no dialog featured all three developers. Seven threads featured one developer, and four threads featured two developers. The developers appeared to be distributing their efforts to respond to user comments, resulting in 67% coverage of all the various discussion threads initiated by users. This division of labor, reducing the effort required from each individual developer, is promising when considered in light of the previous observation that developer participation is crucial for collaborative dialogs to occur.

7. DISCUSSION

Earlier in this paper a conceptual distinction was established between cooperation and collaboration, that pivoted about the role that coordination played in each. Following Dillenbourg, cooperation, we argued, involved coordination of structure, such as in the division of subtasks that might only be fully integrated at the end of a project. Collaboration, on the other hand, involved coordination of process, in which participants sought on an ongoing basis to work together on a common project. The data from JESSE and UMADA suggested that in these review forums this distinction exists in productive tension.

Collaboration. We saw widely different levels of collaborative dialog across the two cases, with relatively little occurring in JESSE and with a lot of short, focused dialog between developers and reviewers occurring in UMADA. The UMADA dialogs do appear to be generating much useful formative input to improve resource design. The differences between these cases could be explained by the fact that the UMADA community is a mature one with a long history of working together. JESSE is a new journal and has yet to form any kind of community identity. It thus appears that the first order design goal of supporting collaborative dialog between participants is not always met in the Partnership Review Model; the issues may be more related to shared cultural values or historical factors rather than technology or process design.

Cooperation. In both cases, we observed interesting cooperative behaviors. The interface affordances of the environment facilitated participants' monitoring of the actions of others and adjustment of their own activities accordingly. In JESSE, this resulted in complementary review coverage across the components of the educational resource made visible in the outline provided in the review environment. This outline essentially provided discussion participants with a summary of the complex resource, making visible its overall structure and some of the key interactive components. Reviewers could therefore not only engage directly with the units that they think they might be interested in, but they could also see which of the units were being reviewed by their peers. Participants explicitly referred to this fact; they reported seeing the ongoing review activities to be both motivating and rewarding. They also commented on its perceived value in reducing their individual workload. In the case of UMADA, the discussion among reviewers also showed cooperative behavior, particularly in the generation of new design requirements.

Preparing Resources for Review. It has been discussed how the Partnership Review Model was adopted by two different settings in response to specific design requirements. The anchored discussion forum, featuring a shared representation of the resource structure, was thought to be an appropriate tool for the peer review of complex educational objects for several reasons. Preparing the object for review with the D3E Publisher's Toolkit created useful and functional representations of the objects under review. While the Toolkit can automatically parse HTML resources reliably, it is not designed to parse resources implemented in other programming or mark-up languages, or resources containing multiple HTML pages. In the case of the Ice Age Module, its multiple HTML pages had to be merged before processing with the Toolkit. While the Toolkit's final output closely mirrored the original structure of the Ice Age Module, its navigational features were lost. This design issue, of how to preserve and represent the interactive qualities of a resource, is one that does not crop up when designing peer review environments for scholarly texts. It would be impossible for the Toolkit to support every language used to create educational resources, however it would be beneficial to extend the Toolkit to support a wider variety of commonly used languages and to support the parsing of related sets of HTML pages.

This highlights the implications of a key design decision when using the Partnership Review Model: whether to use the object itself to create the anchored environment, or whether to use a representation of the object. In the case of JESSE, the editors decided to use the Ice Age object itself, which resulted in the processing complications just discussed. In the UMADA case, the developers chose to use a representation of the object, in the form of the software documentation. This made the processing of resources easier, but created extra work in that the documentation had to be written. It did, however, enable intangible properties of the resource to be included in the structural outline, such as '3D

navigation metaphor.' Recent research by Churchill et al. have demonstrated the use of a tool called Anchored Conversations, which attaches a chat window to documents in native Windows applications like Word, PowerPoint and Photoshop.

8. CONCLUSION: SAYING AND SEEING IN THE PARTNERSHIP REVIEW MODEL

In the two cases reported here, it appears that the utility of an anchored discussion forum for the reviewers of complex educational objects lies not only in what it allowed reviewers to *say*, but in what it allowed them to *see*. Inter-reviewer cooperation, in this interpretation, appeared to consist not so much of the creation of consensus through dialog, but rather in agreeing to allocate tasks in certain ways. This latter form of agreement was facilitated by the review environments' interfaces, which displayed the educational objects split into discrete units for discussion and review. Using the interface, the units of an educational object could not only be 'adopted' by various reviewers according to their particular fields of expertise, but they could also be *seen* to be adopted by all the reviewers. In both cases, the review environment thus facilitated a visual partition of the object under review that emphasized the key structural elements of the educational object, a visualization that in turn appeared to prompt reviewers to distribute the reviewing effort among them according to whoever felt expert in a particular area.

Reviewing these observations, it is suggested that what appears to be happening in these cases is that the review environments' interfaces are surfacing the complex structures of the review objects in ways that promote cooperative behavior among reviewers. The analysis of reviewer interactions in terms of task coverage and completion, and also in terms of (lack of) duplicated effort, suggested that reviewers were cooperating, either intentionally or unintentionally, to divide and manage the task among them. These findings suggest that *cooperative* behavior among participants in these review forums is therefore an emergent, second order property of the technology that exists alongside its primary, first order property of reviewer collaboration.

These findings speak to existing peer review research in several ways. They run counter to analyses of the peer review process reported earlier that emphasize the facilitation of dialogic communication between reviewers as one way to overcome inter-reviewer difference. They also exist in tension with some models of computer mediated peer review that also see improved inter-reviewer communication—this time computer mediated—as a key foundation of successful peer review.

The cooperative behavior observed in JESSE and UMADA proved functionally useful for the review of large-scale complex educational objects where no individual has the time or expertise to provide coverage for the whole object, and indeed where time is probably at a premium for all concerned. This latter point has particular implications for the DLESE model of essentially voluntary community-supported digital library development, in which community participation is to be encouraged. The Partnership Review Model thus appears to offer good potential for informing the design of community-based review of complex educational objects. It leverages the interdisciplinary nature of the reviewer community against the heterogeneity of educational objects they will be reviewing. It does this by enabling the distribution of reviewer expertise across disciplinary aspects of educational objects in ways that are not dictated by the editors or moderators of the discussion, but which emerge out of the individual expertise that each reviewer brings to the review.

9. ACKNOWLEDGEMENTS

The research described here was partially funded under Cooperative Agreement #ATM-9732665 between the National Science Foundation and the University Corporation for Atmospheric Research, and the NSF Grant #0085600.

10. REFERENCES

[1] AAAS Benchmarks for Science Literacy. American Association for the Advancement of Science (AAAS), Washington DC, 1993.

[2] Archer, J. Electronic publishing in the humanities and social sciences, http://ahdp.lib.ucalgary.ca/hssfc, 2000.

[3] Arms, W.Y. Digital Libraries. MIT Press, Cambridge, 2000.

[4] Axelrod, R. The Evolution of Cooperation. Basic Books, New York, 1984.

[5] Buckingham Shum, S., and Sumner, T. JIME: An Interactive Journal for Interactive Media. First Monday, 6 (2 February), 2001.

[6] Churchill, E.F., Trevor, J., Bly, S., Nelson, L. and Cubranic, D., Anchored Conversations. Chatting in the Context of a Document, in CHI 2000, (2000), ACM Press, pp. 454-461.

[7] Cicchetti, D. Reliability of reviews for the American Psychologist. Am. Psychologist, 35. 300-303, 1980.

[8] Cicchetti, D. The reliabilty of peer review for manuscript and grant submissions. Behavioral and Brain Sciences, 14. 119-186, 1991.

[9] Cicchetti, D. and Conn, H. A statistical analysis of reviewer agreement and bias in evaluating medical abstracts. Yale Jnl. of Biology and Medicine, 45. 373-383, 1976.

[10] Cockburn, A. Writing Effective Use Cases. Addison-Wesley, Menlo Park, 2001.

[11] Dillenbourg, P., Baker, M., Blaye, A. and O'Malley, C. The Evolution of Research on Collaborative Learning, in Reimann, P. and Spada, H. eds. Learning in humans and machines. Towards an interdisciplinary learning science, Pergamon, London, 1995.

[12] Fischer, G. The Importance of Models in Making Complex Systems Comprehensible, in Tauber, M. and Ackermann, D. eds. Mental Models and Human-Computer Interaction 2, Elsevier Science Publishers B. V., North Holland, 1991.

[13] Fjällbrant, N. Scholarly communication: Historical development and new possibilities, http://educate.lib.chalmers.se/IATUL/proceedcontents/fullpaper/nfpaper.html,, 1997.

[14] Hargens, L. and Herting, J. A new approach to referees' assessments of manuscripts. Social Science Research, 19. 1-16, 1990.

[15] Harnad, S. Peer commentary on Peer Review. Behavioral and Brain Sciences, 5 (2). 185-186, 1982.

[16] Harnad, S. Scholarly skywriting and the prepublication continuum of scientific inquiry. Psychological Science, 1. 342-343, 1990.

[17] Hollan, J., Hutchins, E. and Kirsh, D. Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction, 7 (2 (June)). 174-196, 2001.

[18] Hutchins, E. Cognition in the Wild. MIT Press, Cambridge, MA, 1996.

[19] Kastens, K.A. How to identify the "best" resources for the reviewed collection of the Digital Library for Earth System

Education. Column "Another Node on the interNet," Computers & Geosciences, 27 (3). 375-378, 2001.
[20] Khoo, M. Community design of DLESE's collections review policy: A technological frames analysis, in ACM/IEEE JCDL 2001, 2001, 157-164.
[21] Khoo, M., Ethnography, evaluation, and design as integrated strategies: A case study from the WES collection, in 5th European Conference on Digital Libraries (ECDL 2001), (Darmstadt, Germany (Sept. 4 - 8), 2001), Springer, 263-274.
[22] Kuhn, T. The structure of scientific revolutions. University of Chicago Press, Chicago, 1970.
[23] Manduca, C.A. and Mogk, D.W. The Digital Library for Earth System Education: A Community Plan. Final Report from the University of Oklahoma to the NSF, Grant 99-06648, 2000, 44.
[24] Marlino, M.R., Sumner, T., Fulker, D.W., Manduca, C.A. and Mogk, D.W. The Digital Library for Earth System Education: Building community, building the library. Communications of the Association for Computing Machinery (ACM), 44 (5, May), 2001.
[25] McKie, D. The scientific periodical from 1665 to 1798. Philosophical Magazine, Commemoration Issue. 122-132, 1948.
[26] Orlikowski, W. and Gash, D. Technological Frames: Making Sense of Information Technology in Organizations. ACM Transactions on Information Systems, 12 (2). 174-207, 1994.
[27] Peters, D. and Ceci, S. Peer-review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Sciences, 5 (2). 187-195, 1982.
[28] Psycoloquy. http://www.princeton.edu/~harnad/psyc.html.
[29] Putnam, R.D. Bowling alone: America's Declining Social Capital. Journal of Democracy, 6. 65-78, 1995.
[30] Riggs, T. and Wilensky, R. An algorithm for automated rating of reviewers. ACM/IEEE JCDL 2001. 381-387, 2001.
[31] Siemens, R. The credibility of electronic publishing, http://web.mala.bc.ca/hssfc/Final/Credibility.htm,, 2001.
[32] Speck, B. Publication peer review: An annotated bibliography. Greenwood, Westport, CN, 1993.
[33] Speier, C., Palmer, J., Wren, D. and Hahn, S. Faculty perceptions of electronic journals as scholarly communication. JASIS, 50 (6). 537-543, 1999.
[34] Sumner, T. and Buckingham Shum, S., From Documents to Discourse: Shifting Conceptions of Scholarly Publishing, in Human Factors in Computing Systems (CHI '98), (Los Angeles (April 18-23), 1998), ACM Press, 95-102.
[35] Sumner, T., Buckingham Shum, S., Wright, M., Bonnardel, N., Piolat, A. and Chevalier, A., Redisigning the peer review process: A developmental theory-in-action, in COOP 2000: Fourth International Conference on the Design of Cooperative Systems, (Sophia Antipolis, France, 2000).
[36] Weatherley, J. Adapting educational resources for collaborative on-line peer review, MS thesis, Dept. of Computer Science, University of Colorado, Boulder, 2001, 87 pages.
[37] Zuckerman, H. and Merton, R. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. Minerva, IX (1). 66-100, 1971.