SKEN: The Science Knowledge and Education Network

Building a User base around Scientific Publications: Editing Online Content and Annotating Scientific Materials.

Steve Kelling, stk2@cornell.edu
Paul Allen, pea1@cornell.edu
Rick Bonney, reb5@cornell.edu
Alan Poole, afp7@cornell.edu

Cornell Lab of Ornithology, Ithaca NY

PROJECT SUMMARY

As part of the NSDL request for services to increase the impact, reach, efficiency, and value of the digital library, we propose to develop an open-source infrastructure to create a knowledge and education network—a new and powerful application for building dynamic collaborative communities centered around primary scientific references. We call this application a Scientific Knowledge and Education Network (SKEN), which will transform the primary scientific references into "living" publications that include the most current information on their topics, and which allow continuous annotations of the content through community input.

Intellectual Merits: SKEN will allow access to huge amounts of peer-reviewed scientific information. However, it also will allow users, including members of the public, to contribute new information on an ongoing basis. And, it will facilitate harvesting of data and resources from other parts of the Internet. This new information will then be vetted and edited by scholarly authors and editors before being uploaded to the primary content. Finally, educators will be able to use SKEN to locate information for their teaching, as well as to post teaching resources and strategies back to the community. The underlying goals of SKEN are to expand traditional, content-based scientific information into a community-based information exchange, and to provide an innovative mechanism for blending science knowledge with opportunities for formal and informal science education.

Broader Impacts: We believe that SKEN will significantly advance current methodologies of research dissemination and validation. Through the use of advanced Information Technologies and the Internet, SKEN will move primary scientific resources from restricted content distribution points (bricks and mortar libraries) to ubiquitous availability over the Internet. It also will decrease the time required to update these scientific resources, and will provide significantly easier access, search, and archival capabilities. By ensuring that all scientific communities implementing SKEN architecture become part of the NSDL infrastructure, we will provide easy cross-fertilization of scientific disciplines.

Finally, the SKEN organization and dissemination of primary scientific resources, and development of the digital library tools that will allow easy manipulation of those resources, will meet major needs in public education. SKEN will allow significant opportunities to discover, explore, and integrate information in previously unimaginable ways. Our audience will include not only the scientists, researchers, authors, and editors who create the content, but also thousands of teachers who use SKEN's organizing tools to integrate content directly into their curricula, along with tens of thousands of students who will have access to original scientific information in their course of studies.

SKEN also will help bring scientific information to audiences, including minorities and students in less wealthy areas, who might otherwise have limited access to primary scientific content.

STATEMENT OF NEED AND PROJECT GOAL

Every field of science, from astronomy to zoology, gains data and knowledge each day, and the results of scientific studies can be eclipsed in weeks.

Happily, the Internet has revolutionized the ease with which traditionally published information can be disseminated. Researchers who once relied on paper journals and photocopy machines can now access scores of scientific journals with the click of a mouse. For example, a subscription to BioOne (www.bioone.org) provides digital and searchable access to the papers published in 73 journals focused on the biological sciences. In the physical sciences, ChemPort (www.chemport.org) provides access to articles from 5,000 electronic and 200 paper journals focused on all aspects of chemistry.

Yet the Internet can go much further in nurturing science as a living and growing body of knowledge. Beyond "electronic publishing," which simply moves paper publication to an electronic platform, newly evolving resources in Information Technology (IT) can lead to a global "knowledge network" (Ginsparg, 2001). As part of the NSDL request for services to increase the impact, reach, efficiency, and value of the digital library, we propose to create the infrastructure for developing a Science Knowledge and Education Network (SKEN). This new and powerful open-source infrastructure will provide a dynamic and interactive platform for knowledge dissemination, and will be useable by essentially any scientific discipline. SKEN will enable collaborative communities to be built around primary scientific references, transforming them into “living” publications that include the most current information on their topics, and which allow continuous annotations of the content through community input. As proof of concept, we will use the SKEN infrastructure to create new digital communities around 3 primary scientific references: Cornell Lab of Ornithology’s Birds of North America and Home Study Course in Bird Biology, and the Smithsonian Institution’s Encyclopedia of Life.

Important steps toward developing such a network already have been taken. For example, many encyclopedias, such as Access Science, the McGraw Hill Encyclopedia of Science and Technology (www.accessscience.com), are now found online. Online encyclopedias can be updated much more easily than paper versions. Books, too, can be published online, allowing authors to keep contents current. An example is The Arabidopsis Book (www.aspb.org/publications/arabidopsis) recently published by the American Society of Plant Biologists. Or consider *UpToDate*, a subscription-based medical reference whose entire contents are reviewed and revised by physician authors and editors each quarter. Published evidence is summarized and specific recommendations are made for patient care (www.uptodate.com).

Even more exciting are websites that facilitate scholarly discussion along with the publishing of scholarly information. An example is Cognet, a MIT Press website that includes access to journals, references, and books on brain and cognitive science research. Cognet strives to create a community of interacting neuroscience professionals by posting news, announcements, seminars, and calls for papers (cognet.mit.edu).

Equally interesting are sites built entirely through public participation, such as Wikipedia (en.wikipedia.org), a collaboratively produced open-content encyclopedia. Anyone can add information to the site or edit existing information. Wikipedia is self regulated, that is, information is contributed and validated by community consent. Therefore, content may or may not contain the most accurate or up-to-date information.

Another innovation is Connexions (cnx.rice.edu), an open-source, open-content, community-driven approach to authoring, teaching, and learning. This online educational environment allows authors or communities of authors to develop course modules within XML frameworks that reside in a content commons. Educators can repurpose and aggregate these modules into larger units that can be grouped into chapters and texts. Connexions also opens the editorial process to third-party reviewers, and allows users to preferentially locate and view modules, chapters, or texts of interest.

The goal of SKEN is to blend and add to the above-mentioned approaches and functionalities to create a new open-source infrastructure for scientific and educational collaboration that will be available at no charge. SKEN will be developed to conform to the open access standards of the National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) initiative. Like earlier applications, SKEN will allow access to vast amounts of peer-reviewed scientific information, but it also will allow users, including members of the public, to contribute new information on an ongoing basis. And it will facilitate harvesting of data and resources from other parts of the Internet. This new information will then be vetted and edited by scholarly authors and editors before being uploaded to the primary content. Finally, educators will be able to use SKEN to locate information for their teaching, as well as to post teaching resources and strategies back to the community. SKEN will thus expand traditional, content-based scientific information into a community-based information exchange, and provide an innovative mechanism for blending science knowledge and opportunities with formal and informal science education.

Demonstration Case

To develop, demonstrate, and test our proposed application infrastructure we will use *The Birds of North America* (BNA), a recently published, hard copy, 18-volume series containing life histories of our continent's breeding birds. This definitive work, which includes one article (species account) for each of North America's 716 species of nesting birds, results from a decade-long collaboration among more than 700 volunteer authors and reviewers—mostly professional ornithologists—in the US and Canada. BNA received its financial and scientific backing from three major North American organizations: the American Ornithologists' Union (AOU; www.aou.org); the Cornell Laboratory of Ornithology (CLO; birds.cornell.edu); and the Academy of Natural Sciences of Philadelphia (ANSP; www.acnatsci.org). Released in print to 2,200 subscribers and considerable acclaim, BNA has recently been acquired by CLO and is being prepared for online publication.

BNA provides ideal content for developing SKEN. It includes over 18,000 pages of authoritative information; an established network of volunteer authors and reviewers who remain committed to the project; name recognition; a large and established subscriber base; and a subject matter that reaches the heart of the continent's citizen science community. More than 2 million individuals in the US are believed to be serious birdwatchers, and thousands of them contribute data and sightings to burgeoning numbers of online databases (e.g., www.birds.cornell.edu/LabPrograms/CitSci/index.html), illustrating how successfully the enthusiasm and expertise of the North American birdwatching community can be harnessed for scientific ends.

We propose to use the rich contents and established community of BNA developers and users as the prototype for developing SKEN. We will build an application infrastructure to convert BNA from an authoritative but static publication into a dynamic knowledge network that continuously receives, edits, and publishes new findings from ornithologists, educators, and citizen scientists across the continent. When this infrastructure is complete we will make it available as an open-source content management and dissemination application, providing other scientific disciplines the opportunity to manage primary scientific content and to build interactive communities around the tasks of increasing, improving, understanding, and teaching about primary scientific resources.

Many tasks will be involved in developing the architecture for SKEN. These include 1) building a content publishing and editing mechanism to allow a) authors to create new or update existing content and b) editors and peer reviewers to modify content; 2) developing a metadata infrastructure to allow finer-grained organization of content held within the primary content; 3) creating and releasing content harvesters that a) search the Internet for new and pertinent content and b) provide it to users; 4) producing the infrastructure to develop online discussion forums to further information dissemination; and 5) creating digital library tools that allow searching, annotation, and creation of educational materials based on primary scientific content. Finally, we must create an OAI-compliant portal between SKEN and the NSDL. Each of these tasks is described below, and our intention is to release SKEN using BNA as its demonstration in 2006. A detailed timetable and checklist (see final sections) will be used to monitor our progress in meeting this aspiration.

TARGET AUDIENCE

The communities that will use dynamic and interactive scientific publications comprise many types of individuals: 1) authors, editors, and reviewers of the primary scientific content, or those who might write or edit articles related to the content; 2) research scientists, who will scour the primary content to gain new information on topics they are studying, 3) faculty and students, such as those who will use the content to prepare lectures or reports, and 4) members of the public who wish to deepen their understanding of the topic.

Authors, Editors, Reviewers: These individuals will benefit from SKEN in numerous ways. First, we envision a more efficient processing of manuscripts submitted for publication through SKEN. Web-based review management tools will provide an online forum where authors, editors, and reviewers can simultaneously access common material to discuss changes, additions, the relevancy of new material, and more. Providing common access to material under review is a clear advantage to the editorial process; also, content that is web based can be instantaneously updated and published once new material and changes are approved.

In addition, authors require rapid access to current literature, sorted in categories that aid synthesis and writing. Our applications will comb existing literature, archive information by taxonomy and subject, and post findings to a common site where authors can access the information quickly and easily, and where editors can view the same material when critiquing an author's work. Most established search engines (e.g., Google) are not effective in assembling literature reviews because they find excessive peripheral material. The SKEN search engine will focus only on the best of the current literature, and will provide a more focused review of that material.

Equally important, our application will provide access to information not yet published in refereed literature, including data that might not ever be published. Such information could include new findings from established researchers and their students, data from organized citizen science projects, or information from interested members of the public. Such information can have great value to authors writing or maintaining primary content and to editors updating or improving existing content because it can fill gaps in information, and because it often provides a leading edge to key findings that follow. When opinions differ regarding the accuracy or value of particular information or data, these views can be posted as well. SKEN will thus help with the current problem that many solid and useful research findings are never published, because the scientist making the findings judges them irrelevant to a focused study, or because the scientist decides that other findings take precedence for publication, or because the findings occur outside the science mainstream. Authors may even find important new ways to use previously unavailable data as they consider where the information fits into the larger picture or a different context. We thus foresee great synergy as new, unpublished findings from related fields become more readily available to authors and outside researchers.

Research Scientists: Researchers will find that SKEN provides "one-stop shopping" for primary scientific information in their field, including key findings and the literature that supports those findings. Our application will boost the value of primary content tremendously by providing the latest information in a researcher's own and related fields, combed from the published literature; by promoting synergy and cross-fertilization among researchers; by opening doors into information that is inaccessible via traditional sources; and by posting and vetting new data quickly, circumventing the time-lags inevitable in traditional publishing.

Teachers and Students: Large repositories of content, such as BNA with its 18,000 printed pages of material, can be overwhelming for students and teachers. SKEN will

contain a suite of digital library tools to allow rapid access to primary content. It also will include abstracting templates that help users select and create new information presentations quickly and succinctly.

These features will benefit educators in several ways. For starters, teachers will be able to prepare reading lists, or to select specific topics and annotate material for class discussions. For example, an ecology teacher might wish to use BNA to examine bird physiology in relation to habitat. Using abstracting templates, the teacher could create a specific presentation of BNA content that pulls together the physiology and habitat sections of four exemplar species, allowing students to see the contrasts among them.

SKEN will also provide many benefits for students, including quick research on key subjects, the ability to browse up-to-date scholarly information for research topics, and the chance to participate in discussion groups with a variety of experts and citizen scientists, where they will gain exposure to multiple viewpoints and differences of opinions. In addition, students involved in research will have the opportunity to contribute to a scholarly resource. The informal setting we envision for SKEN will encourage such contributions, while its editing processes will help students develop rigor in their scientific approach and writing as their contributions are critiqued by reviewers and editors.

Public: SKEN will also provide value to the interested public. First, because of its open access, SKEN will provide all citizens with rapid access to accurate and up-to-date scientific information, which they will know has been checked and edited by experts. In addition, individuals will be able to watch science at work as they read about and even contribute to ongoing science investigations. This will help to create a sense of community among scientists and the public. Thus SKEN will take the concept of citizen science, by which members of the public contribute data to ongoing research projects, to a whole new level in which citizens can track the fate of the data they helped to collect.

PROJECT DESIGN AND OBJECTIVES

Above we identified the six major objectives to make SKEN useful and effective. To reiterate, we must first develop the information architecture and online tools to allow authorized individuals to publish and edit the primary content. Second, we must develop a metadata infrastructure to publish content through existing digital library infrastructures, and allow the population of metadata frameworks for the primary content. Third, we must identify and integrate the appropriate application software for discussion forums. Fourth, we must develop Internet search routines that will seek out relevant information on predetermined topics and integrate this information with the primary content. Fifth, we must create a shareable library of digital teaching resources. Finally, we must integrate SKEN into the NSDL framework and provide a seamless connection between all of its information resources. These six objectives are elaborated below.

Publishing and Editing Primary Content: The foundation of SKEN will be the primary content, one or several peer reviewed information resources that serve as the nexus for information sharing. To remain relevant this content must evolve with the state of knowledge in the field. Therefore we will develop an online content-management infrastructure that allows authors to create content or to edit existing primary content, and allows editors to ensure that changes conform to prescribed publication standards. This infrastructure will have appropriate rights management functionality to control misuse of assets. Updates, modifications, and new information could include new images, figures, tables, and rich media (audio and video), as well as text.

In the case of BNA, the primary content is the 716 original species accounts, which already exist in digital (XML) structures. Once these accounts are loaded into our content management system, the cycle of online publication and content updating will begin. Authors will be able to access their accounts, add information, edit accounts with web-enabled tools, and release updates to editors. Editors will then be able to approve changes, make their own changes, or discuss editorial suggestions with authors. New account versions will become available when an editor approves updates. This process will use many standard features of a content management system (CMS) including:

- separation of content and presentation so that authors deal strictly with content
- integration of multiple authors and editors to prevent conflicting content changes
- management of tables, graphs, bibliographic data, and rich media
- no requirement for users to have knowledge of HTML or XML
- document versioning and automatic comparison of differences
- flexible workflow to accommodate interactions among authors and editorial staff
- security to protect content and audit trails to record changes.

We are focusing our attention on Plone (plone.org), a mature open-source CMS built on an open-source content-management framework called Zope (www.zope.org). Both Zope and Plone have large user communities and are actively undergoing development, so we believe that both packages have a bright future. Several high profile sites are already using Plone, including NASA (www.nasa.gov) and the Mars Rover Maestro project (mars.telascience.org). Plone has also been successfully used to implement the collaborative education Connexions Project.

The CMS requirements outlined above are available through Plone via a web browser interface. Built-in tools allow authors and other content creators to focus on content, largely without worrying about presentation. Plone also allows authors and editors to give editing rights to specific content pieces to facilitate collaboration. It provides simple edit-review workflow by default, and also has a richer workflow available through simple customization.

Plone has several, rather generic, types of content that it handles “out-of-the-box.” It handles basic documents (web page content), files, folders, events, images, topics, links, and news items, and allows users to create and manage these objects. Additionally, Plone/Zope provides a way to create new content types and plug them into the CMS. In fact, a Zope package called CMFBibliography implements a content type for

bibliography entries. To address the requirements for BNA we will create additional content types that correspond to individual BNA accounts and their components. We will also create content types that correspond to contextual metadata (see below).

Plone contains a WYSIWYG content editor, thus content creators require no knowledge of HTML or XML. It also provides an undo feature so that changes can be rolled back. Additionally, it comes with security and roles ensuring that only authorized users can create, edit, review, delete, and otherwise manage content.

Developing a Metadata Infrastructure: Information extraction is easiest when content is organized by predefined data types and/or coded in a way that enables a user to locate, modify, and retrieve particular text components. Most digitized scientific content has technical metadata attributes that define its resource location, version, ownership, and provenance information. To ensure maximal interoperability with external portals and information harvesting mechanisms, our technical metadata will conform to standards established by the NSDL (metamanagement.comm.nsdlib.org/outline.html). While Plone/Zope natively supports the Dublin Core metadata standard, but there is no built-in publishing capability in Plone/Zope that conforms with the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH; www.openarchives.org/OAI/openarchivesprotocol.html). In developing SKEN, we plan to create a Plone/Zope module that will implement a service to publish the metadata for the content of a Plone/Zope installation conforming to OAI-PMH. Provided to the Plone/Zope community, this would be a major step in making existing content in Plone/Zope installations available to the world via the NSDL initiative.

Less common in scientific literature are contextual metadata attributes, which specify how to represent objects, concepts, and other entities within a discipline. The creation of these ontologies is under way in a variety of fields. For example, the Access to Biological Collections Databases data exchange schema was developed to organize the contents of natural history museum specimens (www.bgbm.org/TDWG/CODATA/). Following this lead, we will develop the necessary infrastructure to allow the creation of contextual metadata within primary content. For the BNA an expert committee, comprising ornithologists and digital librarians, will convene to specify contextual metadata definitions. After the metadata schema is constructed, we will organize authors and volunteer sub-editors to populate it. Contextual metadata can then be used as criteria in searches or included in results that searches produce. Editors, authors, and experts will use a web interface to view primary content as they create or edit contextual metadata for their area of expertise.

The Connexions Project uses Plone to store and manage educational modules as XML documents using CNXML (cnx.rice.edu/technology/cnxml/0.5/spec). Module creators and editors download the XML document for a module and edit it using Authentic, a free product from Altova (altova.com). Users editing the content then upload the edited XML document back into Plone. The Authentic XML editor uses templates created by one of Altova's commercial products (e.g., Stylevision) to provide WYSIWYG editing of XML files whose structure is defined by particular XML DTDs or Schemas. We will be able to

use this same approach by expressing the contextual metadata as an XML document. Authors responsible for populating contextual metadata for primary content could then download the metadata document, use Authentic to edit it (which guarantees that the XML document will remain valid according to the metadata definition), and upload the finished product.

Developing Discussion Forums: To facilitate communication and build a user community around primary scientific content we will employ discussion forums. These forums will not only foster interchange among interested parties, but also will be a key access node to information that might not make it into the primary literature, e.g., in situations where a researcher may be focused on one subject but also may have additional or peripheral unpublished data on others. Such forums will foster synergy among people focused on different aspects or disciplines and will provide an interface between experts and amateurs, or educators and students, with clear benefits for all.

Discussion forums will address topics that cut across the organization of the primary content. (For discussion and annotation of the primary content itself and discussion of single-species issues see the Digital Library section.) For our BNA example, the primary content is organized taxonomically, while the discussion forums (as opposed to annotation) may benefit from organization along other axes of interest (e.g., ecology, physiology, life history). The Plone CMS contains built-in support to create and manage discussion forums. All community members, whether author, editor, educator, or student, will be able to initiate discussions and be full participants in discussions. The forums will be unmoderated, but authors or editors will be able to remove inappropriate posts.

Releasing Content Harvesters: As research continues on any given topic, newly published information is added to the wealth of existing information. For anyone interested in a topic, the capability to monitor new research and collect references is crucial. Bibliographies are a key resource for any discipline, and keeping them updated would be an important community service. Therefore, a key component of SKEN will be automatic harvesting of information from other online repositories, such as *BioOne*. Content harvesting will become part of the auxiliary material associated with the primary content, until the author incorporates it. Since the harvesting is very focused, little mediation of the content will be necessary. Examples of materials that could be targeted for harvesting include newly published articles from select journals, the specimen information from museum collections, and relevant audio or video recordings.

We will harvest bibliography entries using OAI-PMH. Harvesting will target journals of interest (for BNA, examples are ornithological journals, physiology journals, ecology journals, and nutrition journals). Once bibliographic metadata are harvested they will be stored locally and offsite links to web-accessible versions of the material will be created. Bibliographic data and links will be available in species-specific contexts. Integration of OAI harvesters with Plone/Zope should be straightforward since an open-source implementation of an OAI-PMH client (zope.org/Members/infrae/oaipmh) already exists in Python, the language used by Plone/Zope. The methodologies that we will employ

specifically harvest the technical metadata of new content, and not the primary content itself.

Creating a Shareable Digital Library of Teaching Resources: All scientific disciplines offer tremendous potential for educators to create a wealth of digital teaching resources based on primary content. One of our SKEN's primary goals is to provide a platform of content and services that dramatically improves the ability of educators to create and exchange educational materials based on the primary content. Our ability to accomplish this will be through the contextual metadata that the community develops for its primary content, and the digital library services available in SKEN.

Initially the digital library services will include tools for annotating the primary content, comparing and contrasting relationships across resources, and abridging text across accounts to create new content representations. For example, an ecology instructor may wish to use the BNA to contrast the diet and parental care among species of raptors. To do this she would use SKEN tools to create an abridging template that selects the raptor accounts and filters each one, retrieving only the sections on diet and parental care. Next she would save the template and make it public for her students to read.

We will develop our digital library services using the Plone CMS. One of its built-in features gives authors and editors the option to allow comments and annotations to be added to specific sections of primary content. This capability allows readers to directly annotate and discuss the primary content, in context, and is an important, probably unique, expansion of traditional online publishing. This built-in feature could be used in the version of primary content available to the public and also as a way for editors and reviewers to provide feedback to authors updating their primary content.

In addition, Plone offers features that provide each user with tools and a workspace to create their own content or to abstract existing content, and thus has huge potential for enhancing primary content. Educators will be able to upload course documents and additional materials that tie into primary content. Researchers will be able to share datasets and preliminary results, and students can share research papers. Both educators and researchers will be able to search the primary content and select pertinent information that can be formatted into new information presentations. These content abstractions can then be made generally available. Also, if editors feel that new "secondary" content is of general interest and provides additional value to the primary content, they can link the secondary content to the primary content.

Moreover, we will integrate the Connexions project module concept and its XML specification for educational content CNXML (cnx.rice.edu/technology/cnxml/0.5/spec/) with the iLumina Digital Library project (www.ilumina-dlib.org) method of sharing materials created by educators. Neither of these projects provide direct access to primary content; for example, Connexions assumes that module creators already have access to primary content, while iLumina provides links to primary resources. To go beyond this, SKEN will integrate the education tools directly into primary resource materials, thus

providing the means to create materials, a central repository to store them, the tools to find them, and the context for those materials (i.e., the primary content).

Finally, we will develop search routines to enhance access to the primary content. We will provide key word searches and will use heuristics and other rules to provide relevancy rankings of search results. Most importantly, we will expand the built-in search capabilities of Plone to be aware of the contextual metadata content and to allow search utilities to be developed around the mining of metadata content. For example, searches of BNA content will use several different metadata sets to provide powerful search functionality. Most users will already be familiar with searching the technical metadata (e.g., Dublin Core) of the accounts and will expect to be able to perform targeted searches typical of library systems (e.g., by title or author). Users also will be able to search for species accounts that match criteria applied against specific primary contextual metadata (e.g., all accounts that mention "Alaska" in the "Distribution" topic). Once secondary contextual metadata are refined and populated, users will be able to perform very targeted searches (e.g., search for accounts of species that have body mass less than 50g).

The contextual metadata will allow users to specify exactly what they want to see as the results of their search. For example, instead of the entire content of matching accounts, users could specify subsets of metadata from matching accounts. Thus metadata could be used as a source of entirely new datasets for analysis. Finally, users will be able to save searches (e.g., search criteria and specifications for results markup) and share them with others in SKEN.

Plone/Zope has built-in search capabilities for free text search of content as well as targeted searching of specific content types. One of SKEN's significant contributions will be to enhance this standard search capability, and to handle searches of specific metadata fields beyond text matching and Dublin Core. In this way searches could be made on numerical or date comparisons. This will enable finely detailed searches of metadata to compile results that contain very specific results.

Integration into the NSDL framework: Qualified Dublin Core metadata and contextual metadata for BNA/SKEN content will be made available to NSDL and other harvesters via implementation of OAI-PMH as described above.

KEY STAFF

Key staff for this two-year collaborative project are divided into four groups:
1. Information Technology and Application Development: PI Steve Kelling (Director of the Information Technology [IT] Program at CLO) has over 10 years of experience in managing and developing interactive Internet applications, and was recently granted a patent for one of them (Kelling et. al 2004). Co-PI Paul Allen is the lead application programmer for CLO's IT program and has many years of experience both in programming for the Internet as well as research in ornithology. Along with half of Allen's salary, this proposal also seeks a fulltime salary for two years for an assistant application developer. Together, the developers will ensure that the multiple tasks and

application integrations outlined in this proposal are accomplished. The application team will work closely with CLO's award-winning web design and information architecture team to ensure that the user interfaces remain at the high standards already in place. The proposal requests funds to cover 20% of one designer's salary for both project years.

2. BNA Test Case—User Community, Primary Content Revision, and Production: BNA provides excellent primary content for use in building SKEN. In addition to its 1,800 pages of text, it boasts a robust group of subscribers –over 2,200, about half libraries and half individuals—who are committed users of BNA's content. Initial marketing surveys suggest a strong interest in an online BNA, especially with continual updating and the addition of sound and video.

Co-PI Dr. Alan Poole has been editor of BNA since its inception (1991), and brings more than a decade of experience with BNA's content and the extensive ornithological community that produced it. As chief editor of BNA Online, Poole will be responsible for content-management and revisions, as well as for the coordination and continued mobilization of the BNA community. BNA authors, over 700 strong, volunteered their time to create the species accounts for the series, and they remain fully committed to keeping the material up to date by serving as ongoing reviewers and sub-editors. In short, BNA is a strong and viable project with an extraordinary volunteer community dedicated to its persistence.

3. Education: Co-PI Rick Bonney, director of education at CLO, has over 25 years of experience in producing educational materials about biology and natural history for nearly all audiences. Founder of CLO's citizen science program, he has received nine awards from the National Science Foundation's directorate of education and human resources to develop curricula, teacher education programs, citizen science projects, and exhibits focused on birds. He also is an editor of CLO's Home Study Course in Bird Biology, the most extensive general ornithology text currently in print. Bonney will coordinate the educational aspects of SKEN, including integration into CLO's existing curriculum development project, Classroom BirdWatch (ESI-0242666). This proposal also seeks funding for a full-time education assistant who will help Bonney with many tasks: collecting front-end data on user requirements, particularly the needs of educators who will use SKEN; conducting usability testing on web interfaces; interacting with educators to answer questions and help them learn to use the network; writing user manuals; and conducting a final evaluation to ensure that SKEN has met its initial goals.

4. Consultants and Affiliates: Given the multiple components of this project and the diverse clientele that SKEN intends to serve, we have recruited a number of consultants to work with us. Dr. William Arms (Professor, Dept. of Computer Sciences, Cornell University) has and will continue to provide input on the overall integration of our project with NSDL infrastructure (see letter of support), and we shall continue to rely on Diane Hillmann and John Saylor, both of NSDL, for critiques of our project and metadata issues. Dr. Jack Bradbury, Director of the Macaulay Library at CLO, will advise on the integration of rich media (sound and video) into SKEN, particularly focusing on issues of access control and rights management (see letter of support). Teresa Ehling, Director of

Electronic Publishing at the Cornell University Library, will assist us in our efforts to develop digital publishing mechanisms (see letter of support). Finally, we will work closely with Dr. Jason Mobley, director of CLO's Home Study Course in Bird Biology, and Dr. Terry Erwin, of Smithsonian Institution's Museum of Natural History to publish their projects through SKEN (see letters of support).

TIMELINE

We believe that within 18 months of project onset we will have BNA fully functioning within a SKEN environment, and that 6 months thereafter we will have SKEN available for use by other projects. Consequently, we believe that we can accomplish these goals by the end of 2 years, and we provide the following timeline as a guide:

Phase	Tasks	Completion Date
I	Install Plone and Zope CMS on CLO server; load BNA XML into Plone; develop rough GUI for viewing BNA accounts through Plone; Customize Plone to provide authoring and editing tools for primary content of BNA; make BNA accounts Internet accessible.	6 months after project initiation
II	Create OAI-PMH interface and implement technical metadata for BNA; convene workshop to develop contextual metadata for BNA, and begin metadata creation.	12 months
III	Refine GUI for BNA; customize Plone for topic-specific discussion forums, and provide annotation capabilities within each account; release suite of digital library tools including search tools and abridging templates; develop and release authoring and teaching tools; release a fully functional BNA within SKEN environment.	18 months
IV	Develop generic GUI functionality for SKEN; expand functionality used in BNA to function for other communities; develop another knowledge network through SKEN for the Home Study Course of Bird Biology or Smithsonian's Encyclopedia of Life.	24 months

DISSEMINATION AND BROAD IMPACTS

An online framework that can organize, interpret, and disseminate the vast quantity of primary content now available in the sciences through focused communities of authors, editors, researchers, educators, students, and the interested public will have tremendous scientific and social benefit. SKEN will profoundly impact the way that primary science content is accessed, updated, and used for research and education purposes. Disseminating scientific knowledge to a broad audience and allowing that audience to

interact, annotate, and create new resources from the knowledge base will help scientists, educators, and students address significant fundamental issues within their disciplines and learn about and better understand those disciplines.

The SKEN software that we develop will be available at no cost under an open-source license. We believe that SKEN will significantly advance current methodologies of research dissemination and validation. Through the use of advanced Information Technologies and the Internet, SKEN will move primary scientific resources away from restricted content distribution points (bricks and mortar libraries) to a ubiquitous availability over the Internet. It also will decrease the time required to update these scientific resources, and will provide significantly easier access, search, and archival capabilities. By ensuring that all scientific communities that implement SKEN architecture become part of the NSDL infrastructure, we will provide easy cross-fertilization of scientific disciplines.

Finally, our organization and dissemination of primary scientific resources, and the digital library tools that will allow easy manipulation of those resources, will meet major needs in public education. SKEN will allow significant opportunities to discover, explore, and integrate information in previously unimaginable ways. Our audience will include not only the scientists, researchers, authors, and editors who create the content, but also thousands of teachers who use SKEN's organizing tools to integrate the content directly into their curricula, and tens of thousands of students who will have access to original scientific information in their course of studies.

CLO has significant experience in disseminating educational materials to a wide audience. Largely through previous NSF support, we have created a robust education and citizen science program that reaches students of all ages and ethnicities in both formal and informal settings. We have placed special emphasis on urban and minority audiences through development of Project PigeonWatch (ESI-9802248) and Urban Bird Studies (ESI-9802248), on teachers through our Educator's Guide to Bird Study (www.birds.cornell.edu/schoolyard; ESI-9618945), and students through Classroom FeederWatch (ESI-9550541) and Classroom BirdWatch (ESI-0242666). SKEN will continue these emphases and also will help to involve women in the sciences. Already women comprise more than one-third (37%) of the existing BNA author base, and many of the women authors are younger ornithologists. We will continue to build on this base and will seek to improve this percentage, actively recruiting women as revising authors, reviewers, and sub-editors.

Online communities are inclusive by their nature, but we are committed to ensuring that SKEN is inclusive by targeting both women's and minority colleges and universities to make them aware of the educational resources available through SKEN, starting with BNA. Nearly all colleges and universities have ecology and zoology programs; many are not aware of BNA and the wealth of information that it provides. We see the SKEN framework as an excellent way of reaching out to smaller, less well-endowed universities. We believe that the education modules that SKEN will develop will be especially useful

to smaller colleges that often have trouble raising the money and expertise to develop solid science curricula.

EVALUATION

The goal of SKEN is to create a new open-source infrastructure for scientific and educational collaboration that can be used for a variety of scientific disciplines and communities. The SKEN concept will be tested by building a collaborative, online version of BNA. Evaluation will thus assess the degree to which SKEN is successful in making BNA useful to an online audience of authors, editors, researchers, students, and teachers. If SKEN can meet this need on behalf of BNA, it should be useful for other audiences and disciplines as well.

Evaluation of SKEN will include front-end evaluation to assess user requirements; formative evaluation to determine whether the network is clearly organized and easy to use; and summative evaluation to determine whether SKEN is able to develop an online community such as the one we envision.

In front-end evaluation we will survey representatives of all user groups to determine what features and functions each group desires in an online collaboration. Information will be gathered by email and through a small number of focus groups, most likely conducted in association with professional meetings such as the annual meetings of the American Ornithologists' Union or the National Science Teacher's Association.

Then, as SKEN is being developed, we will conduct ongoing formative evaluation, primarily usability testing with web interfaces, including those for inputting and editing information; for locating information within SKEN; and for the abstracting templates used by educators. Usability testing will occur with individuals or small groups and will employ goal-based scenarios in which users will be tracked as they attempt to perform pre-set tasks using the web interfaces. CLO is experienced in usability testing through development of its many citizen science projects, including The Birdhouse Network, eBird, and the Great Backyard Bird Count. Additional formative data will be solicited through email and online questionnaires.

After BNA online has been developed using the SKEN application, summative evaluation will determine the satisfaction of each user group in accessing, organizing, and using BNA content. Through online questionnaires and telephone calls, we will measure the extent to which the user requirements, set during the front-end evaluation, have been met by the network. We also will assess the extent to which researchers and authors are actually using SKEN in their daily work. If SKEN is shown to meet the needs of its various audiences, we will publish the evaluation results to encourage other disciplines to adopt SKEN for building new scientific communities online.

Evaluation will be coordinated by Rick Bonney, SKEN co-PI and director of education at CLO, and carried out by the project's education assistant in consultation with other members of the CLO education team who are experienced with evaluation

procedures. Bonney has coordinated evaluation of all CLO's citizen science projects and also is collaborating with Jack Bradbury, director of CLO's Macaulay Library, in evaluating his current NSDL grant (872612445).

SUSTAINABILITY

The Cornell Lab of Ornithology (CLO), a department of Cornell University, is also a nonprofit membership organization. Both membership subscriptions and Cornell University support will help to provide long-term sustainability for the infrastructure (both personnel and hardware) of our proposed project.

While we will make the SKEN software and documentation freely available for downloading and installation under an open source license, CLO will provide fee-based consulting and services to other organizations that do not have the technical expertise to install, customize, and support their own SKEN site.