Cyberinfrastructure for International, Collaborative Biodiversity and Ecological Informatics

A Report of the Pan-American Advanced Institute (PASI)

Held at the Organization of Tropical Studies (OTS)

La Selva Biological Station in Costa Rica,

on May 31 through June 12, 2008

James H. Beach
Biodiversity Research Center
University of Kansas
Lawrence, Kansas

Julio E. Ibarra
Center for Internet Augmented Research and Assessment
Florida International University
Miami, Florida

19 June 2009

Funded by the U.S. National Science Foundation

Office of International Science and Engineering, Award #0617469 and #0617616 to Florida International University and to the University of Kansas.

Table of Contents

Table to Tables

Executive Summary

Research on biological diversity is a global and cyberinfrastructure-enabled science. The evolutionary origin of biological diversity and ecological processes that maintain it are active areas of research involving thousands of scientists worldwide. The outputs of those systematics and ecological investigations inform conservation management, government policy and civil society on the most important environmental phenomena in the history of mankind—the impacts resulting from the transformation of species natural systems into species-poor, managed ones.

While profound changes are affecting earth's biosphere and climate, networked cyberinfrastructure is changing the pace and methods of biodiversity research. Global- and continental-scale environmental issues demand accelerated research strategies, and rapid progress requires timely, global- and continental-scale collaboration. International collaboration with biodiversity and ecological informatics will build the common network protocols for taking on 21st century environmental science challenges. Biodiversity informatics training will be critical to develop the proficiencies needed to build and sustain that shared cyberinfrastructure.

A Pan-American Advanced Studies Institute (PASI) was held at the Organization of Tropical Studies (OTS) La Selva Biological Station in Costa Rica, on May 31 through June 12, 2008, to address these challenges. The PASI emphasized the application of cyberinfrastructure tools for environmental research collaboration. The PASI served advanced graduate students and young faculty from North, South and Central America by teaching the application of software tools and services to biodiversity and ecological research problems through lectures, demonstrations and hands-on lab tasks and rain forest data collection exercises.

The goals of the PASI were to (1) expose biodiversity research and biodiversity informatics students from the U.S. and Central America to advanced concepts in distributed network-based science enabled by cyberinfrastructure tools, (2) to bring together students and faculty from biology and computer science and engineering disciplines to share knowledge and explore new approaches to domain science problems through the use of CI, and (3) create an environment that sensitized the students to exchange ideas and consider new approaches to research through the use of CI and (4) for that information to persist in a repository for future use.


Blue Jean frog

The curriculum and structured program were central to satisfying the goals of the PASI. The curriculum provided a broad set of topics to stimulate exchange between students in the biodiversity and ecology disciplines with students from the computer

science and engineering disciplines. The PASI curriculum focused on the following themes that combined biodiversity science with cyberinfrastructure issues, and provided diverse and rich content to the students about how biodiversity research is transforming from the adoption of a cyberinfrastructure-based paradigm:

- **Biodiversity Data Conceptualization and Management**
- **Scientific Workflows, Data Integration, Semantic Mediation**
- **Biological Collections Data Description, Discovery, Integration and Retrieval**
- **Grid Introduction and Fundamentals**
- **Web Services and Network Resource Frameworks**
- **Environmental Sensor Data Management**
- **Environmental Niche Modeling and Macroecological Analysis**
- **Biodiversity Conservation and the Application of Biodiversity Informatics**

Descriptions for each of these themes, the PASI curriculum, and descriptions of other goals and their outcomes are included in the report sections.

Through the student recruitment process over 100 applications were received. The student selection process resulted in the selection of 28 participants from 9 countries. The distribution of student participants consisted of 9 females, 19 males; 15 participants from Latin America (8 countries), and 13 participants from the U.S.; 19 were from the biology research discipline, 9 from the computer science/engineering discipline. The distribution of instructors who participated were 7 females, 22 males; 11 instructors from Latin America (4 countries), 18 from the U.S.; 9 from the biology research discipline, 20 from the technology research domain.

The U.S. National Science Foundation Office of International Science and Engineering funded the PASI under awards #0617469 and #0617616 to Florida International University and to the University of Kansas, respectively. Partners and local host for the PASI were the University of Costa Rica, a major contributor to the leadership role of Costa Rica in biodiversity research; INBio, the National Biodiversity Institute of Costa Rica, is a research institution that has been internationally recognized as a pioneer in Biodiversity Informatics; the Organization of Tropical Studies (OTS), a consortium of over 60 universities and research institutes from throughout the United States, Latin America, South Africa and Australia, and facilitator of the La Selva field station.

1. Introduction

1.1 Biological Diversity Research

Biological diversity, or simply biodiversity, is the sum of life on Earth↓plants, animals and microbes↓encompassing all levels of biological organization from genomes to species to ecosystems. Approximately 1.8 million species are known as a result of 300 years of the biological exploration of the planet. Astonishingly, an estimated 15−50 million species await discovery and basic description.

A grand challenge for the 21st century science is to harness knowledge of Earth's biological diversity and how it shapes the global environmental systems on which all of life depends. This knowledge is critical to science and society for rational policy for managing natural systems, sustaining human health, maintaining economic stability, and improving the quality of human life. The urgency for this knowledge increases daily as the conversion of natural systems to human-managed systems accelerates the decline of biological diversity.

The importance of biodiversity research and education has been established by a series of landmark reports: U.S. NSF's Task Force on Global Biodiversity (Black et al., 1989), the systematics and biological collections community's Systematics Agenda 2000 (1994), the Australian government's The Darwin Declaration (Environment Australia, 1998), and the U. S. President's Committee on Science and Technology's Teaming with Life (Lane, 1998).

Research in ecology, systematics, biogeography and phylogenetics has long been appreciated as global activity that transcends regional or national boundaries. In the western hemisphere, studies of biological diversity have been international exercises since the great naturalists from Europe explored the new world in the 18th century. Today, much international collaboration exists between North and Central American researchers in the form of surveys and inventories, floras, faunas, and with taxonomic monographic and revisionary studies.

An NSF-funded workshop (NSF/OISE #0549443) held in Panama in January, 2006 brought together biodiversity scientists, network engineers, research policy-makers, and funding agency representatives from throughout the western hemisphere to examine the ways in which international collaborations among biodiversity researchers could be improved through Cyberinfrastructure (CI).

1.2 Rationale for Cyberinfrastructure for International Biodiversity Research Collaboration

Cyberinfrastructure, such as research community computation and collaboration architectures, grid-based services, web services, automated semantic mediation for data integration, standard metadata profiles, standard information retrieval protocols and web-based research analysis applications, are transforming the nature of much biodiversity and ecological research in the 21st century (Foster, 2005). The global nature of

environmental issues and the international distribution of species and researchers demand regional, continental and ultimately global collaboration for biodiversity informatics training. Cyberinfrastructure can mediate much biodiversity research and monitoring collaborations, if the professional capacity is developed among researchers in many countries.

Central America was recommended as an excellent regional target for joint U.S.-Latin America training in biodiversity cyberinfrastructure. There is a strong historical scientific community of research collaboration between Central American and U.S. universities and biodiversity research centers. The University of Costa Rica's longest formal relationship with any U.S. university for international training of students is with the University of Kansas. Finally, the PIs conducted an interdisciplinary workshop on the topic of CI for biodiversity research collaborations in Panama for this region (NSF/OISE #0549443). Findings from the workshop provided recommendations for the development of training and outreach in the use of Cyberinfrastructure tools to enhance the support of collaborative biodiversity research among Central American nations, the U.S., and other countries.

1.3 Next steps from the Panama workshop on Cyberinfrastructure-enabled Biodiversity

One of the recommendations from the Panama workshop was to conduct a Pan-American Advanced Studies Institute (PASI) in Central America. The goal of a PASI would be to give students and young faculty from North, South, and Central America the opportunity to learn about CI tools and their application to biodiversity and ecological research through two weeks of lectures, hands-on technical laboratory exercises, and group discussion. The "Cyberinfrastructure for International, Collaborative Biodiversity and Ecological Informatics" PASI was held in June 2008, in Costa Rica. The PASI would facilitate activities by which to disseminate both scientific and engineering knowledge, and stimulate collaborative learning and cooperation among the research communities of the Americas.

James Beach, University of Kansas, and Julio Ibarra, Florida International University, are the PIs of the PASI. Beach, Assistant Director of Informatics at the Kansas University Biodiversity Research Center and an active biodiversity informatics researcher, was co-lead for organizing the biodiversity research and bioinformatics component of the PASI. Ibarra, Assistant Vice President for Technology Augmented Research of the Center for Internet Augmented Research and Assessment (CIARA) at FIU and recipient of the NSF International Research Network Connections (IRNC) grant award for Latin America, was co-lead for organizing the cyberinfrastructure component of the PASI. The remainder of this report is organized as follows. Section 2 describes the goals, the organization and the setting where the PASI took place. Section 3 discusses the PASI curriculum and the program of activities that combined biodiversity science and CI concepts. Section 4 describes the lecturers and subject matter experts that participated in the PASI, the student selection process, and who were the students that participated. Section 5 explains the outcomes. Then, finally, section 6 contains the appendices with ancillary information.

2. Pan-American Advanced Studies Institute in Cyberinfrastructure for International Collaborative Biodiversity and Ecological Informatics in Costa Rica

2.1 Goals of the PASI

The goals of the PASI were influenced by the findings from the workshop in Panama: Improving International Biodiversity Research Collaboration with Cyberinfrastructure: A Report of the Workshop, "Cyberinfrastructure for International Biodiversity Research Collaboration". In addition, the PIs engaged participants of the workshop, as well as domain experts to provide insights towards the development of the curriculum, its content, and programmatic activities to achieve the goals of the PASI.

The goals of the PASI were to (1) expose biodiversity research and biodiversity informatics students from the U.S. and Central America to advanced concepts in distributed network-based science enabled by cyberinfrastructure tools, (2) to bring together students and faculty from biology and computer science and engineering disciplines to share knowledge and explore new approaches to domain science problems through the use of CI, and (3) create an environment that sensitized the students to exchange ideas and consider new approaches to research through the use of CI and (4) for that information to persist in a repository for future use.

2.2 Organizing Committee

The Organizing Committee (OC) for the PASI was created for the purpose of assisting the Principal Investigators with the organization and implementation processes of the PASI. In particular, the composition of the OC was designed to include domain expertise in biodiversity research, bioinformatics, computer science/engineering and information sciences in order to advise the PIs on the student selection process, recommendations for lecturers and subject matter experts, the curriculum content, and laboratory exercises.

Organizing Committee members were:

James Beach: Assistant Director Informatics, Biodiversity Institute, University of Kansas. Beach is an active collaborator in the field of biodiversity informatics, with numerous software engineering collaborations and leads the Lifemapper and Specify Software projects, two significant ongoing cyberinfrastructure efforts for the biodiversity collections research community.

Eric Graham: Programmer Analyst, Center for Embedded Networked Sensing (CENS), University of California Los Angeles (UCLA). Graham is a national leader in the design, configuration and deployment of mobile environmental sensor network hardware and collaborates with U.S. and international scientists to install terrestrial and aquatic environmental sensor networks.

P. Bryan Heidorn: Associate Professor, Library & Information Science, University of Illinois at Urbana-Champaign. Heidorn is a leading scientist in the field of information

retrieval and natural language processing and works in the biodiversity informatics field. He recently moved to the University of Arizona to head an informatics center.

Julio Ibarra: Assistant Vice President for Technology Augmented Rsearch, Center for Internet Augmented Research and Assessment (CIARA), Florida International University. Ibarra has been involved in the support and management of international networks in the context of enabling international science research and collaboration for 15 years.

Vladimir Lara: Dean of the School of Computer Sciences and Informatics, University of Costa Rica. Lara is an active researcher on the semantic web and the application of ontologies for conceptual mapping. He has co-organized several recent informatics conferences in Central America.

Gabriela Marín Raventós: Vice-Dean, Graduate Studies, Universidad de Costa Rica. Marin is a researcher and leading organizer of informatics and computer science conferences in Latin America. She has interests in supporting the role of women in Computer Science and the role of collaboration technologies for enabling science.

Erick Mata: Director of Biodiversity Informatics, Instituto Nacional de Biodiversidad (INBio). Mata is a computer scientist and known for his research, training and cyberinfrastructure contributions to the field of biodiversity informatics. He is co-Chair of the GBIF science committee and a frequent contributor to international collaborations.

Robert "Bob" Morris: Professor of Computer Science, University of Massachusetts, Boston. Morris is a highly-respected, computer scientist and mathematician, and an active researcher in ecological and biodiversity informatics for 15 years. His research has made numerous significant contributions to international standards and related conceptualization in those fields.

Deana Pennington: Research Assistant Professor, LTER Network Office, University of New Mexico. Pennington is a computer scientist with a extraordinary strength in human learning, postgraduate training events, and workshop and institute organization. With NSF CI-Team funding, she leads an effort to bring new geospatial processing technologies to early-career biologists.

2.3 Partners and Local Hosts

Costa Rica was selected as the preferred host country, because of its leadership role with biodiversity research, inventory and informatics in Central America, and because of its higher education relationships with U.S. institutions. Our partners in Costa Rica for the development and implementation of the PASI were the University of Costa Rica (UCR), INBio and the Organization for Tropical Studies (OTS).

The University of Costa Rica (UCR) is a major contributor to the leadership role of Costa Rica in biodiversity research. Scientists at the UCR have developed important collaborative working relationships with colleagues at universities in the U.S., in education and research of tropical biology. UCR provided critical guidance in local logistics, the student selection process and the contribution of students to the PASI.

INBio, the National Biodiversity Institute of Costa Rica (www.inbio.ac.cr), is a research institution that has been internationally recognized as a pioneer in Biodiversity Informatics. It conducts an exhaustive inventory of Costa Rican biodiversity that has led to the development of sophisticated software tools to capture, manage and disseminate digital information about Costa Rican biodiversity. INBio representation in the Organizing Committee, the student and faculty selection process, the curriculum development and their participation in the PASI was a pivotal contribution.

The Organization of Tropical Studies (OTS) is a consortium of over 60 universities and research institutes from throughout the United States, Latin America, South Africa and Australia, whose aim is to strengthen education, research and rational use of natural resources in the tropics. OTS owns three biological stations in Costa Rica and offers several highly regarded graduate and undergraduate-level courses each year in tropical biology. The La Selva biological research station, located in northern Costa Rica, is one of the world's most productive biodiversity research sites.

2.4 Setting for the PASI

The setting and activities of the PASI were critical elements towards achieving these goals. Considerable time and effort was contributed by the co-organizers, the local hosts, and the organizing committee to identify a location that could provide the essential facilities for a lecture series, while also providing a real-world environment for field work and diverse motivating laboratory exercises. La Selva was an ideal site for the PASI, offering excellent facilities for the curriculum and its laboratory exercises, combining the use of CI tools in a real-world biodiversity environment, http://www.ots.ac.cr/index.php?option=com_content&task=view&id=162&Itemid=348

Rufous-tailed Hummingbird

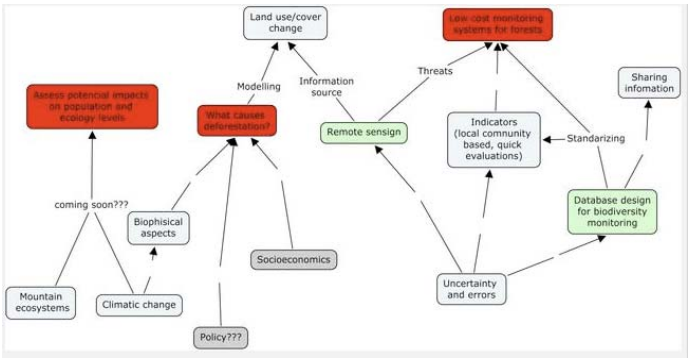
La Selva is also the site of the implementation of a state-of-the-art environmental sensor network (Sensor Arrays) that will be used to characterize the 4-dimensional environment of the lowland tropical rainforest to enable many kinds of sustained observations that are currently unachievable. The instrumentation that uses the environmental sensor networks at La Selva will be used to collect data to address questions of global importance, and it will assuredly stimulate innovative lines of sensor-enabled research. Providing students with the opportunity to witness the development of a state-of-the-art environmental sensor network allowed the PASI organizers to recognize that La Selva was an ideal venue for the PASI.

2.5 Learning Technologies for the PASI

One of the main objectives of the PASI was to expose students to Cyberinfrastructure (CI) tools that incorporate new computational and network methods being used in biodiversity and ecological research. The curriculum (Appendix, section 6.1) was

designed to include real world technological tools for students both to witness how they are used, as well as to use the tools to perform in class and field laboratory exercises.

A networked environment was created in the lecture/learning facility. Each student had a laptop, either their own or a loaner. Servers were installed with the software that was used for laboratory exercises, as well as to create a repository where students could store and retrieve data during the course. The wiki and repository were left available for students and faculty to continue using it after the PASI, to be preserved by the Organization of Tropical Studies (OTS). The URL for the PASI wiki and repository is http://www.ots.ac.cr/wiki/index.php/Main_Page.



Most curriculum topics that demonstrated the use of CI tools, included a hands-on session where students gained the experience of downloading the software from the PASI server or from the Web to then install and configure it, to then use it for a laboratory exercise. For example, to understand the concepts of Cyberinfrastructure, students were exposed to concept mapping. They learned fundamentals of concept mapping using the CMap Tools from http://cmap.ihmc.us/conceptmap.html. The overview on Geospatial data and standards, included a hands on laboratory session using the Quantum GIS (QGIS) geographic information system, which is an application for data viewing, editing and analysis (http://en.wikipedia.org/wiki/Qgis.

The curriculum that is included in section 6.1 identifies othersoftware applications used.

3. Curriculum and Program

The PASI curriculum was designed to expose early-career faculty, postdoctorals and advanced graduate students with backgrounds in biodiversity and ecological research to new computational and networked methods. Our goal was to have the participants discover possibilities to leverage their research activities to be more international, networked and collaborative. Biodiversity informatics is a young and multi-disciplinary domain and we deliberately assembled a student/faculty body with varied research interests and technical skills to represent as much of the field as possible.

The PASI organizing committee met over several months in conference calls at 2-3 week intervals. With the original group of volunteers, it was not easy to congeal on a common curriculum. Some committee members wanted the PASI to be highly focused on one or two themes reflecting their own research interests; one committee member withdrew when the consensus was not to take that narrow path. It was also a challenge to get some members of the original organizing committee to participate in the work of curriculum

planning. We ultimately succeeded however and recognized that the best organizers were those who taught or trained students as academicians. The final members of the organizing committee listed in this report proved to be heroic contributors of their time and expertise; each stepped up to take on planning tasks and they all contributed significantly to the curriculum development and delivery on site. Identifying scientists who enjoy organizing, mentoring and collaboration is an important determinant of success.

Our schedule emphasized interactive sessions in a workshop atmosphere, which alternated with formal presentations, Q&A periods, discussions, hands-on exercises, tutorials and forays into the rain forest for data acquisition. Our daily sessions typically started at 8 AM and extended until 8 PM with breaks for lunch, dinner and coffees (See Section 6.1). Frequently, evening sessions continued until 10 PM, with student presentations or other supplementary material. Overall the pace of the PASI was intensive and exhilarating. Situating the PASI at a field station with facilities designed for research and for meeting was a very good decision. In our view field stations are ideal choices for PASIs as they have few distractions. We had helpful and competent OTS staff supporting our technical and meeting needs.

By mixing up the different kinds of activities during the day we kept the participants active and engaged. Communications and dialogue among all participants improved quickly once the context was set and the members of the PASI became acquainted with each other's research interests. Satisfaction with the curriculum and pace was high (see student feedback below), although there is clearly no way in two-weeks time to avoid the trade-off between breadth and depth for any particular topic. We opted for breadth in most cases, as most topics to be handled comprehensively would have required two or more weeks each on their own. Given the broadly-scoped and multidisciplinary nature of this young field, and the diverse research interests of those present, the PASI struck a balance between breadth of coverage and depth of focus.

The PASI curriculum focused on several themes that combined biodiversity science with cyberinfrastructure issues, and provided diverse and rich content to the students about how biodiversity research is transforming from the adoption of a cyberinfrastructure-based paradigm.


Howler Money in the trees

The organizing committee selected these themes based on identified hotspots of modern biodiversity and ecological research:

- **Biodiversity Data Conceptualization and Management**

  This theme focused on concepts related to the representation of information from biological systems, at all levels of biological organization, from DNA to species

to ecosystems. Ecological informatics deals with heterogeneous data structures, ranging from spatial data in coordinate systems to grid to graph-structured information representing species interactions to models of community biogeochemical processes. Effective research analysis and synthesis involves dealing with issues of data semantics, representation, spatial and temporal scaling, encoding, time series, transformation, provenance, versioning, and additional metadata concerns. New methods and tools are evolving rapidly, pushed by increasing connectivity and the potential for research integration. (Presentations by: Pennington, LTER; Morris, UMass; Heidorn, Illinois)

- **Scientific Workflows, Data Integration, Semantic Mediation**

  Automated workflow environments are in many ways the ultimate tool for biological researchers as informaticists aim to encapsulate diverse data processing strategies and algorithms into a end-user authoring environments to automate the manual drudgery and breadth of specialized knowledge needed for end-to-end processing of raw data into research insight. Biodiversity and ecological informatics are two disciplines that are test beds for workflow engineering, we demonstrated Kepler, and NSF-funded workflow environment and discussed other tools. The effectiveness of these tools depends completely on the underlying semantic framework to permit the ad-hoc assembly of data sources, transformations, analysis, and output tools. Ontologies, semantic data management and mediation are complex and technically challenging foundations for scientific workflow frameworks. (Thau, UC Davis; Pennington, LTER; Sequeda, Texas)

- **Biological Collections Data Description, Discovery, Integration and Retrieval**

  Species occurrence data, whether observational or museum voucher based, have been aggressively mobilized in the past several years thanks to the development of protocols and interfaces that allow structured data publishing and federated query networks. The biodiversity informatics community has made great progress in establishing community data standards and best practice norms for the acquisition, validation, geo-referencing, and publishing of this data to community data cache's such as the database maintained by the Global Biodiversity Information Facility (GBIF) and the disciplinary networks MANIS, ORNIS, HerpNet, etc. The PASI discussed the concepts, standards and tools associated with managing species occurrence and taxonomic data in a global networked context. (Mata, INBio; Beach, Kansas; Heidorn, Illinois; Shorthouse, EOL/Woods Hole)

- **Grid Introduction and Fundamentals**

  Grid computing is finally starting to show some promise for end user science communities as the concepts and techniques are mapped into workflow tools and interfaces for biological scientists are able to comprehend. Grid computing will surely be vital to high-performance computational analysis and to distributed

collaboration as more biodiversity data become available online and as grid capabilities are more effectively embedded into end user analysis tools. We presented an introduction to grid computing, a hands-on workshop on running simple grid jobs, and discussion about how PASI participants could potentially utilize grid services in their research. (Clifford, Open Science Grid, Chicago).

- **Web Services and Network Resource Frameworks**

  Web services are an increasingly easy way for biodiversity data providers and researchers to integrate applications to take advantage of remote resource availability. Integration frameworks based on web-service protocols are becoming more common. We demonstrated with presentations and with hands-on laboratory exercises examples of web service integration in biodiversity applications. (Shapely, Google; Canhos, CRIA, Stewart, Kansas)

- **Environmental Sensor Data Management**

  Environmental sensor networking is an advancing field in the biological sciences. Advances in sensors, sensor platform and wireless networking technologies have stimulated much recent field research and related engineering. Sensors represent a paradigm-shift for many fields, data driven experiments and long-term monitoring projects are proving to be useful in addressing the pressing ecological questions of our day, such as the physiological response of forest trees to temperature increases, the changing distribution of insects and birds (through audio sampling) with respect to climate change, and the microhabitat characterization of light, energy and gas flux regimes. Sensors are extending the resolution field studies in spatial and temporal dimensions in all directions. The PASI allocated a day on sensor network concepts, technologies, hands-on assembly, deployment and subsequent data analysis. Faculty and staff from UCLA's Center for Embedded Networked Sensing did an exceptional job presenting the intellectual material and the hands-on activities under the direction of Dr. Eric Graham. Dr. Paul Hubbard, UCSD, also discussed middleware concepts of vital importance in sensor network computing.

- **Environmental Niche Modeling and Macroecological Analysis**

  Environmental species niche modeling has recently become a transformative science having been applied to diverse ecological scenarios including the prediction of native ranges to fill in gaps in known distributions, invasive species persistence and spread behavior, the dispersal of infectious disease, impacts of climate change on community species composition, species frequency, parasite and prey relationships, etc. The field is particularly interesting from an informatics perspective because the analytical techniques integrate data across science domains of climatology, species inventory, species dispersal patterns and others ecological dimensions. The conservation planning value of taking archived species occurrence data and applying them to future forecasts of biotic changes due to human mediated climate or land-use changes is enormous. We presented

the fundamentals of niche modeling with a presentation hands-on exercise to illustrate several key data integration, management and analysis issues. (Martinez, UNAM; Stewart, Kansas; Rangel, UConn/Sao Paulo).

- **Biodiversity Conservation and the Application of Biodiversity Informatics**

  The PASI added an element of applied biodiversity informatics with 1.5 days of the curriculum devoted to the real-world processes of biodiversity conservation and the role biodiversity data plays in the conservation management and policies world. The connection of the science we discussed was made clear in the presentations by Dr. Dan Janzen and Dr. Rodrigo Gamez, both accomplished scientists with direct involvement in the national biodiversity planning and national park creation in Costa Rica. These sessions were extraordinary opportunities to review the consequences of biodiversity change and actions that can be taken to mitigate their impact. Both Gamez and Janzen painted the big picture for PASI participants reminding them of the societal values and issues that surround biodiversity research and conservation on a national scale in a hyper-diverse tropical country. Both of these extraordinary scientists graciously volunteered to participate in the PASI without being asked by the organizers.

The curriculum for this PASI can be found in the appendix in section 6.1.

4. Lecturers and Students

4.1 Lecturers

This section recognizes the lecturers of the PASI. The organizers were very fortunate in receiving commitments from these experts to participate in the PASI and to share their knowledge and experience with the students.

PASI People in the forest

Aimee Stewart is senior software engineer from the Biodiversity Institute of the University of Kansas. She is the lead developer on the Lifemapper Project (www.lifemapper.org) and specializes in open source geo-spatial data management and analysis applications. She is a frequent collaborator in community and project workshops and working groups. Stewart lectured and organized a field data collection on topics related to geospatial data management and available software tools for the same.

Anthanasios Stathopoulos a postdoctorate from UCLA along with Dr. **Eric Graham**, Ph.D. candidate **Yeung Lam**, and software engineer **Fabio Silva,** Information Sciences

Institute, University of Southern California, comprised the sensor network faculty for the PASI. This group has unparalleled expertise in the fabrication and development of environmental sensor networks, being a part of the UCLA Center for Embedded Network Sensing (CENS). The group organized an exceptional day-long tutorial on the theory and concepts of sensor networks, the implementation and deployment off sensors in a rain forest environment (in a field exercise), and subsequent data harvesting and analysis using 'R' analytical software. This was one of the training highlights of the PASI.

Ben Clifford, Ph.D., Open Science Grid, University of Chicago, presented an introduction to grid computing and organized a led a hands-on tutorial on defining and running grid jobs.

Bryan Heidorn, Ph.D., University of Illinois, lectured on EML, the Ecological Metadata Language, Darwin Core, and other cyberinfrastructure standards and their relevance and application to the field.

Daniel Janzen, Ph.D., and **Winifred Hallwachs**, Ph.D., University of Pennsylvania organized a day devoted to the ecological informatics of their entomological research in the Area de Conservación Guanacaste (ACG). Janzen also gave an unforgettable two hour presentation on the history of conservation development and biotic surveys and inventories in Costa Rica in the ACG headquarters building in Guanacaste for the group.

Dave Thau, Ph.D. candidate from the University of California, Davis, presented on the use and future of ontologies and semantic mediation and workflows in environmental biology. Thau is a highly-creative researcher leading several notable projects and collaborations in the field, including AntWeb.

David Clark, Ph.D., University of Missouri, St. Louis, gave a presentation on informatics associated with his research on long-term forest measurement and inventory in the rain forest at La Selva.

David Shorthouse, Ph. D., from the Encyclopedia of Life (EOL) Project (Marine Biological Laboratory, Woods Hole, Mass., gave an overview of EOL (www.eol.org) and a hands-on tutorial on the use of Content Management Systems for integrated web-based biodiversity publishing and management.

Deana Pennington, Ph.D., Long Term Ecological Research Network, University of New Mexico, led off the PASI with a a conceptual mapping exercise of the research interests of the participants using the shareware 'C-Map'. Deanna also gave an overview of the biodiversity informatics field and set the context for many interdisciplinary discussions held during the PASI.

Dora Canhos, Ph. D., Centro de Referência em Informação Ambiental (CRIA), Campinas, Sao Paulo, Brazil, provided an overview of CRIA's work and emphasized database and web service applications developed by CRIA to help organize Brasilian collections around a shared data network and services. CRIA has been a leader in the biodiversity informatics field on a global scale.

Enrique Martinez-Meyer, Ph.D., Instituto de Biologia, Universidad Nacional Autonoma de Mexico, presented a session on ecological niche modeling using GARP and other algorithms.

Erick Mata, Ph.D., INBio, Santo Domingo, Costa Rica, provided a historical perspective of biodiversity in Costa Rica, and important international groups (including GBIF and TDWG) working to standardize biodiversity and ecological data management.

James Beach, Ph.D., University of Kansas, was a co-organizer and co-PI of the PASI, and also presented a case study in software engineering based on Specify Software, a museum database application. He included aspects of data modeling, user centered design principles and network integration opportunities for the Specify Project.

Julio Ibarra, Florida International University, PASI co-organizer, presented a lecture on western-hemisphere, international research networks and their application in the support of as virtual research community and collaboration.

Manuel Vargas, Ph. D., INBio, lectured on the concepts and implementation of software supporting the TAPIR information retrieval standard for federating biological collections into a single distributed query network.

Orlando Vargas and **Enrique Castro,** Organization for Tropical Studies, La Selva Biological Station, Costa Rica, organized a half-day session on automated species diagnosis and indentification using online services and databases.

PASI Students, presented 15-20 minute presentations during the course of the two weeks on their own research which was followed by Q&A and discussion in each case.

Paul Hubbard, Ph.D., San Diego Super Computer Center (SDSC), lectured on software architecture and middleware for ecological sensor networks, which included laboratory exercises and field work.

Rebecca Shapley, Ph.D., of Google, lectured on software development and services at Google with application to biodiversity science.

Robert (Bob) Morris, Ph.D., University of Massachusetts at Boston, lectured on data collection and management concepts, ontologies, and the use of a variety of applications.

Rodigo Gamez Lobo, Ph.D., Virologist, ex-Vice Rector for Research, University of Costa Rica, co-Founder, General Director and President of INBio gave the first presentation of the PASI, a stirring, comprehensive review of the history of biodiversity conservation and research in Costa Rica. Dr. Gamez's presentation covered decades of the development of Costa Rica's awareness and protection of its biodiversity resources.

4.2 Student Selection Process

The selection process for the students required that students first complete a 1-page application web form. This form contained a questionnaire to be answered by each applicant. The questions solicited information from the applicant about their research from areas of biodiversity or from computer science. The process allowed the organizing

committee to select a balanced number of biodiversity domain scientists to be paired with computer and engineering scientists.

Student applicants were recruited through directed e-mailings, open dissemination of PASI information in a PDF brochure sent over several mailing lists in ecology and environmental informatics. We also created a web site for the PASI, which contained the text of the original NSF proposal and information about how to apply. The web site is online at: http://www.ciara.fiu.edu/eco. The applicant's faculty advisor was involved as well, providing a formal letter of support for the students.

We received over 100 applicants for the PASI and in the end selected 28 participants from 9 countries. The organizing committee reviewed all of the applications on several criteria and strived for a balance of gender, national origin and research interests. We sought both biology and technology students, as the PASI was interdisciplinary by nature.

The 28 students and 29 instructors were represented in the proportions indicated in the table below.

	Number of Students	Number of Instructors	Total
Female	9	7	16
Male	19	22	41
Latin Americans	15 (8 countries)	11 (4)	26
U.S.	13	18	31
Biology Researchers	19	9	28
Technology Researchers	9	20	29

Table 1 Distribution of Students and Instructors

It was difficult to find and attract female students to the course who were qualified and identifiably interested in the interdisciplinary nature of the PASI. For instructors, technology researchers outnumbered biology researchers over 2:1 because the emphasis of the course was on bringing new computational, hardware and informatics approaches to biodiversity research problems. Students were in the opposite ratio for the same reason.

4.3 Students

The following students were selected to participate in the PASI.

Name	Country	University	Field of Study/Research
Agarwal, Ankit	USA	University of Kansas	Computer Engineering/ Requirements specification for software architecture for a biological research station.
Albuquerque, Andrea	Brazil	National Institute for Amazonian	Biology/ Ontology based

		Research - INPA	semantic data extraction
Arnillas Merino, Carlos Albert	Peru	Universidad Nacional Agraria La Molina	Biology/ Design and implementation of biodiversity monitoring systems
Boyle, Sarah	USA	Arizona State University	Biology/ Effects of forest fragmentation on the spatial distribution centered on tropical ecology and conservation
Brunetta, Bruno	Brazil	Universidade Estadual de Lodrina	Biology/ Ecology and bioacustic of a Brazilian jay
Calder, Matt	USA	University of Massachusetts	Computer Science/ Sensor network reasoning and data inference
Campos dos Santos, Jose Laurindo	Brazil	The National Institute for Amazonian Research	Computer Science/ Researcher in Biogeo Informatics
Dewey, Tanya	USA	University of Michigan Museum of Zoology	Ecology and Evolutionary Biology/ Science education research at both K-12 and undergraduate levels.
Earl, Stevan	USA	Arizona State University	Ecology / Site Manager for the Central Arizona Phoenix Long-Term Ecological Research (CAP LTER) project
Ferrer-Paris, Jose Rafael	Venezuela	Instituto Venesolano de Investigaciones Cientificas	Ecology/ Patterns of species abundance, distribution and species richness
Gomez, Daniela	Argentina	Universidad Nacional de Rio Cuarto	Ecology/ rodent population ecology
Gutierrez, Jaime	Costa Rica	Instituto Tecnologico de Costa Rica	Computer Science/ Analysis, design, implementation and customization of software tools for exchanging information on biodiversity
Lanfri, Sofia	Argentina	National University of Cordoba	Biology/ Modeling of environmental requirements of lizard species in Argentina and construction of predictive distribution maps
Londsdale, Owen	USA	Smithsonian Institution	Biology/ Developing the Biosystematic Database of World Diptera (BDWD)
Luther, David	USA	University of Puerto Rico, Rio Piedras	Ecology/ Automated Remote Biodiversity Monitoring Network (ARBIMON)
Madrigal, Oscar	Costa Rica	Organization for Tropical Studies	Computer Science/ Information Management Area of the Organization for Tropical Studies

Peri, Francesco	USA	University of Massachusetts	Computer Science/ Fields of Wireless Ad Hoc Sensor Networks, Smart Sensor Networks and Semantic Data Validation
Pineda Lizano, Willy	Costa Rica	Tirimbina Rainforest Center - University of Costa Rica	Biology/ Conservation of the tropical rainforests of Sarapiquí, Costa Rica, promote scientific research, and environmental education.
Principe, Romina	Argentina	Universidad Nacional de Rio Cuarto	Ecology/ Ecological effects of small dams on benthic communities from mountain streams
Pulgarin Diaz, John Alexander	Columbia	Universidad Nacional de Colombia	Entomology/ Ecology and systematics of Scolytinae (Coleoptera: Curculionidae) in the neotropics
Rangel, Thiago	USA	University of Conneticut	Ecology/ Understand factors that drive broad scale patterns in species richness, particularly latitudinal gradients in biodiversity
Rodriguez, Josephine	USA	University of Illinois	Ecology/ Microgastrine parasitoid wasps, which attack almost all caterpillars, the dominant herbivores in terrestrial ecosystems
Segura, Angel	Uruguay	Universidad de la Republica	Ecology/ Marine Biology, with emphasis on ecology
Sequeda, Juan	USA	University of Texas at Austin	Computer Science/ Creating a system that will transparently enable existing relational databases to be integrated with the semantic web
Smith, Lori	USA	University of Illinois	Library and Information Science with an emphasis in Data Curation
Solano, Braulio	Costa Rica	Universidad de Costa Rica	Computer Science/ Development software for Information Retrieval (PHP-OpenIsis, a CDS/ISIS open source alternative)
Torres, Juan	Puerto Rico	University of Puerto Rico	Biology/ anthropogenic effects on coastal marine ecosystems, specially coral reefs
Whitman, Melissa	USA	Evergreen State College	Ecological Modeling/Conservation Biology

Table 2 Students selected to participate in the PASI

5. Outcomes

5.1 Cyber-enabled Biodiversity and Ecological Research Repository

A wiki and the Menalto Gallery were installed at La Selva to serve as a repository. A wiki is an excellent tool for the creation and editing web pages in a hierarchical fashion, using a simple editor within the browser, for the management of collaborative work of a community. PASI participants created accounts on the PASI wiki to access information others had placed in the common areas of the wiki, as well as for each person to put their content for others to access it. Within a very short like, the wiki provided a central repository for all the PASI participants to exchange information, text and image format, with others in the course.

The Menalto Gallery is an open source web-based photo album organizer. It uses a gallery metaphor, consisting of a collection of albums containing photos, or any kind of image. Students were encouraged to create their own photo album within the gallery to store photos of animals, insects, classmates or anything of interest while they were at La Selva. By the end of the PASI, the gallery contained many albums with pictures and images participants can return to for research, learning, and to remember the experiences they had at the PASI.

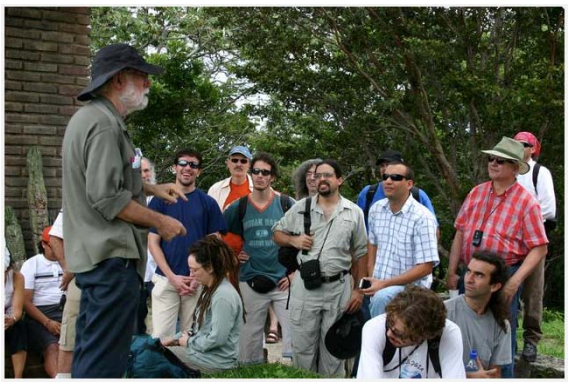
These tools proved to be very useful to retrieve, update and share information among all the participants. The curriculum, lectures, laboratory exercises, presentations, pictures and all other information that was disseminated at the PASI through the wiki and the gallery.

The wiki and the gallery can be accessed at
http://www.ots.ac.cr/wikipasi/index.php/Main_Page

5.2 Course Framework for La Selva

The curriculum was tailored to effectively address the goals of the PASI. The first goal was to expose biodiversity research and biodiversity informatics students from the U.S. and Central America to advanced concepts in distributed network-based science enabled by cyberinfrastructure tools.

The curriculum provided a broad set of topics to stimulate exchange between students in the biodiversity and ecology disciplines with students from the computer science and engineering disciplines. These exchanges sensitized the faculty and students to explore new ways to asking questions about domain science problems through the use of CI. This addressed goal 2.



Dan Jansen explaining his research

La Selva provided an excellent environment to exchange ideas and discover new approaches. The facilities accommodated lectures, with hands on and field work exercises. Social activities created opportunities for additional exchange of ideas, but in a relaxed and open setting. This resulted in several students forming groups to work together during class and laboratory exercises.

The network and servers at La Selva provided the required tools by which all information generated and collected for the PASI was archived. The repository is providing a resource for persistence of this information for current and future use by all of the participants of this PASI, as well as other parties on interest.

The goals of the PASI were to (1), (2) to bring together students and faculty from biology and computer science and engineering disciplines to share knowledge and explore new approaches to domain science problems through the use of CI, and (3) create an environment that sensitized the students to exchange ideas and consider new approaches to research through the use of CI and (4) for that information to persist in a repository for future use.

5.3 Cross-domain knowledge sharing

Exchange and sharing of knowledge across domains was one of the most effective outcomes of the PASI. Students formed groups that consisted of multi-disciplinary members from a domain science and computer and engineering scientists. The domain scientists would present a research problem. The domain scientists and the computer scientists would document the problem to understand its properties and dimension. They would then collaborate to develop a solution that combined domain science methods with CI.

5.4 Post-Course Participant Evaluations

We assessed the quality of the PASI experience for the participants and their level of satisfaction with the event by administering an online questionnaire (www.surveymonkey.com) in the last session of the PASI at La Selva. The complete survey with all responses will be attached to this report in PDF format.

We asked the 28 students 31 questions, some multiple choice and others requiring an open-ended written response. Generally, the level of student satisfaction with the PASI was very high. In the table below the responses to the most important multiple choice questions are tabulated as percentage of the responders applying the answer shown.

(in percent)

Survey Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The PASI was a valuable professional experience for me:	68	30	3	0	0

I found the PASI curriculum to be well-organized:	60	36	0	4	0
La Selva was a good venue for the PASI:	62	23	7	7	0
The PASI sessions offered uniform high-quality content:	46	37	8	8	0
I will follow-up after the PASI with another PASI participant about a potential collaboration:	67	24	Maybe 9		
The PASI had too many, about the right number, or too few instructors:	Too many: 9	Just right: 87	Too few: 4		
The selection of the kinds topics for the PASI was excellent and overall the diversity of topics was a good mix of material	30	52	8	4	0
The mix of students in the PASI was about right.	63	30	3	3	0
The sessions offered valuable and up-to-date information.	50	36	14	0	0
I found the quality of meeting and the overall experience to be enhanced by holding the PASI at a biological field station instead of a meeting hotel or campus.	60	20	12	8	0

Table 3 Participant Table Responses

Participants were asked to identify the top three subjects/presentations they liked best about the PASI, and the top three they liked least:

PASI Topic Rank	Liked Best (number of respondents)	Liked Least (number of respondents)

1	Environmental Sensors and Sensor Networks (11)	Google Presentation (6)
2	Ontologies and Semantic Data Processing (8)	Grid computing (5)
3	Grid Computing (4)	Ontologies (4)
4	Dan Janzen's Presentation on the ACG and Barcoding (4)	Tied: CMAP concept modeling (3), Drupal Installation (3), TAPIR Installation (3), Workflow demonstration (3)
5	Overview of computing and cyberinfrastructure (3)	

Table 4 Ranking of Topics and Presentations

PASI Participants were asked several open-ended questions. This is a summary of the most important of those with excerpted answers chosen to represent typical comments.

Please provide an overall evaluation in 2-3 sentences of the value of the PASI for you and your research interests.

Representative Answers: (All answers were positive; this is a sample of their contents.)

- In this program I had the opportunity to learn a lot in relation to different tools that are going to be useful for my research. I research on stream ecology and I will probably use remote sensing, GIS, spatial analysis and tools to manage databases in my studies.

- It was important for me and my research to know new technologies and software to solve biological problems and the possibilities of collaboration and the interactions between scientists especially when they are from different countries and different funding problems.

- It was an opportunity to interact directly with people with different realities, in the sense of research and minds PASI introduced to me some informatic useful tools to improve or think my research with no computational limitations.

- PASI means a very strongly valuable professional experience for me. Through this PASI I would like to start in Panama a research program group related to most of topics which have been touched during this PASI workshop. Congratulations guys!

- I had an incredible opportunity to view the two sides. As a student from computer science (BioInformatics). I am accustomed to receive data from biologists, but almost ever have no idea of what they pass into the field to produce that data. And the worst is that sometimes they don't even know what they want, and the needs and demands a huge. I have got really impressed with the biologists that came to PASI. All of them were involved with informatics. I made great contacts; we are going to start a project together in collaboration.

- Getting knowledge of new technologies, initiatives and groups working in the same field I am working on, is highly valuable for me and for my institution to keep on working on the projects that have been setup. Also, get contacts with people that have needs and resources that would be necessary to both. I learned many new things that would help to go deep in the conceptualization and finally in the science related to biodiversity usage and conservation.

Question: **What did you like best about the PASI?**

Representative Answers:

- Exciting discussions (formal and informal) about intersections of science and technology in new research areas

- The possibility to interchange research experiences and to learn new technologies that are going to improve our research work.

Collaborating on an experiment

- Wide breadth of topics covered.

- Being at La Selva with this amazing group. Interacting with people in these close situations and learning about their research was where I learned the most and enjoyed the most. Also, Enrique Martinez's presentation was perfect.

- The use of grid technology, seeing all these sources of temporal and spatial data, making new friends, seeing Latin-Americans in a pacific environment, seeing how much Dan Jansen has helped my country.

- I liked being exposed to so many different approaches and tools. I thought the day with Janzen and Hallwachs in Santa Rosa was a highlight.

- The combination between computer scientist and biodiversity people, with many different views of the topics in discussion. The possibility to accommodate the schedule according to the group necessities. The interest of many people in the group to establish communication and cooperation bridges to interchange information and capacity-building.

- Group meetings with collaboration focus. Focusing on merging our research for new applications was excellent.

Question: **What did you like least about the PASI?**

Representative Answers:

- Too long daily sessions sometimes until 8-9:00pm.

- Too much variance in teaching experience among the presenters. Insufficient advanced staging of the software.

- Methodology, I think that some instructors does not have the right methodology to manage the group and present the topic.

- Some technical topics related to informatics that were very difficult to follow.

- Lectures were predominantly discussed from a highly technical and esoteric point of view, and as a result, VERY little could be absorbed by those of us with a biological background – an expected result of mixing highly-specialized scientists from disparate fields.

- The group exercises were not well organized and usually ate up a lot of time that could have used differently.

- Most of the hands on weren't as effective as I thought. I didn't think the Drupal session was necessary. Grid computing was interesting for the CS people, but the biologist didn't get what was going on. Installation of TAPIR was just following a tutorial, and not very useful.

- Unfortunately, some of the presenters appeared unprepared, and did not appear to know their subject material well enough to make their presentation appropriate to the participants.

Question: **Any comments on the organization or logistics of any part of the PASI? Or things that could be changed or improved about the organization of the event?**

- I was amazed on how all the presentations were time-precise.

- I would have included at least a day without lectures to work more on collaborative research.

- Organizers were not sure of the logistics of accommodation/scheduling till the end moment. The travel plans and links between airport and hotel were not worked out until very late.

- The logistics of the PASI were very well coordinated. The schedule was fairly clear and there was flexibility, which is important. It was unfortunate that a day trip to Guanacaste ended up eating up almost 3 days. A side trip that was closer would have resulted in more time for discussions and less down time on a bus.

- Great work with regard to organizing, I can't imagine how much work that must have been.

- I think a reading list before the course to provide some basic information about materials would have been useful. For biologists some readings about databases,

- Make all presentations as equal understandable for CS and Biologist. Many presenters did try to do it and were able to do it. Have a better objective for the hands on, something that the biologist could understand, because everything we did, for us the CS, is straightforward following a tutorial.

- I would suggest fewer lectures. Many of the hour-long lectures could have been summarized quickly, and the participants could have had more time to discuss topics. For example, it was usually adequate to direct us to a web site and explain the tools available. We could then explore it on our own. I also think that the introductions/research presentations would have been more helpful had we done them in the first couple of days. This would have given us an overview of people's research experience and interests. Instead, it took longer to figure out everyone's areas.

Question: **Was the field trip to the Guanacaste Conservation Area a worthwhile experience? And worth the time it took to get there and back? What was the most useful aspect or finding or discovery for you on that trip?**

Representative Answers:

- The possibility of talking to Janzen and share his knowledge is invaluable. The time spent was just a detail. I could hear Janzen directly talking about conservation strategies and his particular experience in ACG.

- Indeed, to spend time in Guanacaste was very good. We had attended a very stimulating talk done by Prof. Dan Jensen and we have got the opportunity to know more about the geological and the environmental plans of Santa Rosa. It was tiring, due the long bus trip, but worth going there.

- That trip was great, Dan Janzen is a legend. Listen to him was for me an honor. No problem how long it took. Unexpected things happen sometime. (Bus breakdown and landslide).

- The trip to Guanacaste was well worth the trip. Seeing the landscape differences, the biological differences, and having a controversial topic was well worth the experience. It broke up the La Selva experience and the workshop in a way that it needed in order to move forward with energy during the last portion. Without the trip, the group would have become restless.

- The most useful aspect was the Janzen talk, because of the knowledge he has. I think that talk generates a lot of conversations between the students and was a very interesting debate.

- Although I had already been to GCA before, I think that the trip was very worthwhile. Spending an entire day with Dan Janzen was worth every second of the bus ride.

- I think that the trip could be a better experience. Guanacaste is in many senses closer to the Latin American experience, so it could be better to take more time there, not only for the "top technology", but for the real and current application of this technology on Latin-American countries.

Question: **Any final questions, suggestions or comments?**

Representative Answers:

- The experience was a great one, I made contact with many possible future collaborators and friends. My most sincere thanks to the organizers for allowing me to participate in such as great workshop! Let's do this again! Thanks Jim, Julio and others!

- A well-run program. Very very valuable and I wish it continued success.

- Dear Julio, Jim, and Bob, you were untiring! Congratulations for your great job in organization, in infrastructure, accommodations, technical/Scientific program. It was a great experience for me. It made me believe that there is still some foundations interested in real science, just science. Your proposal was unite this group for 12 days for discussion and learning and you did, with no charges for us. Do you have any idea of how rare it is? Maybe three new projects to begin, some new technologies presented that must be studied deeply to verify how they could help us, and tones of collaboration opportunities. Great!!!!! Just positive aspects! Please, inform us about the next PASI, I am sure that many of us could be again with you presenting results of some projects conceived here. Thanks you all!!!!!!!!!!

- Thank you!! for the opportunity you gave me to learn, network and meet all these fabulous people!

- Thank you Jim and Julio! This was an experience of a lifetime; I am so grateful and think you both did an outstanding job as organizers.

- This questionnaire was too long. I enjoy very much this workshop; it showed me lots of tools and generates some doubts and plans that will hopefully become resolved and true.

PASI One-Year Later E-mail Survey on Professional Impacts

One year after the completion of the PASI we sent out an email in June 2009 to the participants with these questions:

1. Did the PASI have any long-term impacts on you, perhaps some follow-up activity or e-mail or interactions of any kind that you might have had with the other course participants as a result of the PASI?

2. Did the PASI change anything in your own research plans or program?

From the 28 participants in the PASI (not including instructors) we received 9 substantive responses with these excerpted highlights:

1. The PASI was excellent and you know it influenced the joint work with my advisor Dr Miranker, your group and Greg Riccardi, I have more to say about the influence on my PhD research. I will email you details next week.

2. I tried to develop a project for hexagonal rasters with Matt... but I can't... we advanced and found some information, but I had not enough time to complete it. I keep mail contact with some people for a few months, one for common personal interest, and another one to exchange opinions about biodiversity monitoring. Rafael wrote me to comment a paper. At my office, I am trying to reinforce the development of tools to create web-based databases for biodiversity monitoring, with spatial information (it was not "a change in my life", but a support...and ideas to face some problems). these tools also were a reason to keep contact with Steven. but we did not develop a more formal approach.

3. The course was great in putting together lot of young scientist and senior scientist to discuss very interesting problems. For me it is always great to be in that kind of workshops. It open my mind and charge my batteries to continue on my way. Also hearing (and discussing) Janzen on his place is other of the things that make me be proud of being there. I keep in touch (via mail) with Rafael, from Venezuela, discussing some statistical issues and some free source statistical packages (R) problems. Despite not being in my research area, we share some programs and some discussion with Jaime Gutierrez (INBIO Costa Rica) and John Pulgarin (Colombia). Also we shared information on courses, and get some contacts distributed along America which always is great.

4. In my case the PASI was an excellent opportunity to analyze in a lot of real situations how relationships are between IT specialists and Scientifics. I work in an space where communication between scientists and IT specialists is very important for many reasons, a couple of these: a) The Scientifics need the IT guys to manage easily heir data b) The IT guys need to know how the scientists think to develop to give solutions that meet the scientific need.

5. The PASI definitely impacted my thinking, goals, and research. As far as immediate impacts we collaboratively developed with the Kansas group the successful "Changemapper" proposal as an outcome of brainstorming and discussions at the PASI. I feel my understanding of many aspects of biodiversity research and informatics was greatly expanded and I continue to look for ways to collaborate with the wonderful people I met there and with other members of the

larger biodiversity research community. Also as a direct result of the workshop, the programmer for the Animal Diversity Web (Roger Espinosa) became involved in a Google coding sprint directly related to representing and managing taxonomies for large, biological databases. This also involved the Encyclopedia of Life folks and our interactions with their multiple development teams have intensified over the past year.

6. Well what I have to report should be very obvious to you. Because of the PASI meeting we started our collaboration, which was later formalized by a successfully funded NSF project between UConn and Kansas.

7. Well, talking about the impacts that PASI had on me, I can say that some research projects were started to be discussed on those days and they still go on (towers, connections, sonograms, etc..). The Cyber Infrastructure presented has influenced some of our actions here. Aimee's work guided some of our actions here at INPA. And Dora Canhos is a reference when talking about biodiversity, her work at CRIA is known everywhere. And Juan's work presented in a conference and "spreaded" to us was a great reference. The opportunity of meeting Dan Janzen (the legend, for biologists) and Erick Mata (a reference for us from computer science for biodiversity) was unique.

   No one can deny how PASI influenced my future actions. My Master Degree final work (dissertation) took a brand new direction. David Thau presentation at PASI, observations, and comments on my already published articles made me visualize new horizons for the application of ontology to semantic web in order to promote biodiversity data integration and the use of domain ontology for Amazonian Biodiversity knowledge acquisition. After PASI (understanding much better my dear friends biologists work), I became a kind of specialist on modeling data base/sources on biodiversity domain. MCT (Ministery of Science and Technology) and EMBRAPA has used me as a consultant for these kind of data modeling and this is making a great difference on my academic and professional life. I know what I am doing when talking about biodiversity data modeling, and PASI is part responsible for that. I have left Manaus, traveled to Brasília to help people there.

8. The PASI definitely helped me to branch out a bit more from my world of ecology towards the computer science end of the spectrum. Right now I am in the process of taking a look at interdisciplinary grad programs, trying to narrow down my interests and find a mentor willing to take on a non-traditional student such as myself. I would love to hear any suggestions on who to work with/where to go for a PhD. It is still my goal to incorporate informatics into my future ecological research.

9. The PASI strongly marked my academic history. More than the brainstorms and ideas that came up during PASI (most of them related with bioacoustics), the most

remarkable episodes to me were that afternoon with Dan Janzen and the visit to INBIO. I say that because after our visit to Guanacaste, I was curious to learn about the history of conservation in Costa-Rica and found a lot of interesting and stuff related with the subject that Dan Janzen exposed. Since the PASI I have finished my masters, I have been studying environmental law (taking the Law undergrad here at the University). I am very sorry to disappoint you, once I didn't go to the computers side, but I believe that I will soon be able to contribute more effectively to the world (and Brazil) society and I should thank PASI for this contribution in my academic history.

6. Appendices

6.1 Curriculum and Program

PASI SCHEDULE 17, 12 June 2008

Pan-American Advanced Studies Institute (PASI): Cyberinfrastructure for International, Collaborative, Biodiversity and Ecological Informatics

Schedule of Activities, May 31 - June 13, 2008

Note: Times and Events are likely to change a little during daily schedules at La Selva

Date	Time	Activity	Presenter/Organizer	Comments
Overnight Location				
				Best Western Irazu, Km 3 Autopista General Canas
Saturday, May 31, 2008		International Arrivals to Juan Santamaria Airport (SJO)		San Jose, 1000, CR. Tel 506 2 290 9300
San Jose, Hotel Irazu-		Take a taxi to hotel 18 Km from airport		Pay in CR Colones or US Dollars
Best Western				
	6:00 PM	Informal Group Dinner meet in Hotel Irazu Reception/Lobby at 6:00 PM	All	Walk to Restaurante, La Fuente de Mariscos, 2nd floor, Centro Commercial San Jose 2000, Tel 2231 0631 or 2296 1256
				(Costa Rican participants invited to join us for dinner!)
Sunday, June 01, 2008	7:00 AM	Breakfast and Hotel Check-out		
La Selva Biological Station	8:00 AM	Load Bus at Hotel Irazu with all suitcases		
	8:30 AM	Arrive and settle at INBioparque		
	9:00 AM	**Session 0** Welcome and Presentation on Biodiversity Conservation in Costa Rica	Don Rodrigo Gámez Lobo, President INBio	
	10:00 AM	INBio biodiversity research tour	Erick Mata, INBio	
	11:00 AM	Bus to La Selva		
	12:00 PM	"		
	1:00 PM	Lunch at Rancho Robertos		
	2:00 PM	Back on the bus to La Selva		

		3:00 PM		Arrive La Selva, De-bus, register at reception, pack-in		Checking in, Room assignments, getting settled
		3:30 PM		**Session 1 Facility Tour (optional and if time)**	OTS Staff	
		4:30 PM		Free time (maybe)		
		5:00 PM		Reception Station Director's Patio (mandatory)	Deedra McClearn	Welcome at Family House 1
		6:00 PM		Dinner in the Cafeteria/Comedor		
		7:15 PM		**Session 2 (Overview)** The US NSF PASI Program	Jessica Robin, U.S. NSF	All La Selva PASI sessions are in the PASI conference room in the classroom building, excepting field exercises.
		7:30 PM		Research at La Selva, History, Context, Current	Deedra McClearn	
		8:10 PM		PASI: Introductions, science themes, curriculum and logistics	Beach, Ibarra, Organizing Committee.	Kickoff and Logistics
	Monday, June 02, 2008	6:00 AM		Breakfast (served 6 AM - 7:30 AM daily)		
	La Selva Biological Station	6:30 AM		Optional Guided Tour of Forest (approximately 90 minutes)		Add names on sign up sheets
		8:00 AM		**Session 3 (Overview)** Introduction to Biodiversity Informatics	Mata	
		9:00 AM		Presentation: Conceptual Landscape of Technology Enabled Science	Deanna Pennington	
		10:00 AM		Coffee Break		
		10:30 AM		Hands-on: CMAP software	Pennington/All	Conceptualizing Biodiv and Eco Informatics
		12:00 PM		Lunch (served 11:30 AM - 1 PM daily)		
		1:00 PM		**Session 4 (Tools and Overview)** CMap Hands-On continued		(This serves to introduce everyone's field of
		2:00 PM		Analysis of results of Biodiv-Eco CMaps	Pennington/All	interest and to identify common ground)
		2:30 PM		Coffee Break		
		3:00 PM		Discussion with Individual Research Interests	All	
		4:00 PM		Discussion with Individual Research Interests	All	Journal at the end of the day
		5:00 PM		Free time		
		6:00 PM		Dinner (served 6 PM - 7 PM daily)		

		7:15 PM		**Session 5** Presentation: La Selva: the old-growth tropical rain forest landscape	David Clark, Resident Researcher	
		8:30 PM		Socialize/Free time		
	Tuesday, June 03, 2008	6:00 AM		Breakfast		
	La Selva Biological Station	6:30 AM		Optional Guided Tour of Forest (approximately 90 minutes)		Add names on sign up sheets
		8:00 AM		**Session 6 (Tools):** Specimen, Climate, Geospatial Data, Data Collection/Management Concepts, Issues, Standards	Bryan Heidorn and Bob Morris	Ecological Metadata Language, Darwin Core, SDD, EML, GBIF/TDWG task groups, observations and multimedia
		9:00 AM		"	Bob Morris	
		10:00 AM		Coffee Break		
		10:30 AM		Geospatial Data Overview	Aimee Stewart	OGC standards, Map Server, WMS, WFS
		11:00 AM		Hands-On: QGIS	Stewart and Morris	
		12:00 PM		Lunch		
		1:00 PM		**Session 7 (Tools and Integration)** Hands-On continued	Stewart	specimen data, climate data, geospatial data
		2:00 PM		Lifemapper Introduction		OGC
		2:30 PM		Coffee Break		
		3:00 PM		Web Services using Lifemapper as an example	Stewart	Intro to web services
		4:00 PM		Hands-on: Developing a web services client in Python	"	
		5:00 PM		Free Time		
		6:00 PM		Dinner		
		7:00 PM		**Session 8 (Data)** Geocoding: Birds of a Feather Session	Morris	Geospatial data tools and techniques, TBD.
		8:00 PM				Bring your GPS, camera, cables to connect to a computer
	Wednesday, June 04, 2008	6:00 AM		Breakfast		
	La Selva Biological Station	6:30 AM		Optional Guided Tour of Forest (approximately 90 minutes)		Add names on sign up sheets
				Session 9 (Tools and Integration)		

		8:00 AM	Ecological Sensor Networks	Eric Graham, Yeung Lam		Science Questions, Technologies
		9:00 AM	"	Fabio Silva		Field Exercises Overview & Hardware Setup
		10:00 AM	Coffee Break			
		10:30 AM	ESN's continued	Graham/Lam/Silva		Field Hardware Setup & Deployment
		11:00 AM	Middleware for Sensor Network Communities	Paul Hubbard UCSD		
		12:00 PM	Lunch			
		1:00 PM	**Session 10 (Tools)** ESN's continued	Graham/Lam/Silva		Field Hardware & Data Collection
		2:00 PM	"	Graham		Data analysis for Example Science Questions
		2:30 PM	Coffee Break			
		3:00 PM	"	Graham		Data Analysis, continued
		4:00 PM	"	Lam/Silva		In-depth hardware setup
		5:00 PM	Free time			
		6:00 PM	Dinner			
		7:00 PM	**Session 11 (Tools)** Hands-on: LabView	Lam/Silva		
		8:00 PM	"	"		
	Thursday, June 05, 2008	6:00 AM	Breakfast			
	La Selva Biological Station	6:30 AM	Optional Guided Tour of Forest (approximately 90 minutes)			Add names on sign up sheets
			Session 12 (Community): Grid Computing	Ben Clifford		Grid Computing
		8:00 AM	Grid Computing	Open Science Grid		Overview, Demos
			"	"		Hands-On
		10:00 AM	Coffee Break			
		10:30 AM	Grid Computing (continued)	Ben Clifford		Grid Computing
		11:00 AM	"	"		
		12:00	Lunch			

		PM				
		1:00 PM		"		
		1:30 PM	Unix Tutorial (optional)	Bob Morris		
		2:00 PM	**Session 13 (Community): Grid Computing**	Ben Clifford		
		2:30 PM		"		
		3:00 PM	Coffee Break			
		3:30 PM	Grid Computing Continued	"		
		5:00 PM	Free time			
		5:30 PM	**Session 14: Dinner at Ara Ambigua**	Load buses at 5:30		
		7:00 PM				
		8:00 PM	**Free Time**			
	Friday, June 06, 2008	6:00 AM	Breakfast			
			Optional Guided Tour of Forest (approximately 90 minutes)			Add names on sign up sheets
	La Selva Biological Station	8:00 AM	**Session 15 (Integration): Species Distribution Modeling**	Enrique Martinez		
		9:00 AM	"	Enrique Martinez		
		10:00 AM	Coffee Break			
		10:30 AM	Lifemapper Demonstration	Aimee Stewart		
		11:00 AM	Workflow and Kepler Demo	Deana Pennington and Dave Thau		
		12:00 PM	Lunch			
		1:00 PM	**Session 16: Impact of High-Performance Research and Education Networks on Science**	Julio Ibarra		
		2:00 PM	Coffee Break			
			Session 17 (Integration): Workflow Environments			
		2:30 PM	Integrating your research interests	Deana Pennington		Group discussion - Sensitizing questions

		3:00 PM		Small Groups - using new technologies in your research, Conceptual Workflows Hands on: Paper prototyping workflows	Deana Pennington	
		4:00 PM		"		
		5:00 PM		Free time		
		6:00 PM		Dinner		
		7:00 PM		**Group Presentations: Topics:** GIS and policy, Data Mining, Image Metadata, Models of Collaboration, more ENM, Geo-coding/GIS/ …, other?		
		8:00 PM		"		
		9:00 PM		"		
	Saturday, June 07, 2008	6:00 AM		Breakfast		
	Hotel Boyeros, Liberia	8:00 AM		**Session 18 (Integration): Ontologies**	Dave Thau	
		9:00 AM		Hands-on: Protégé and Ontofest	"	Species profile, SDD, other TDWG ontologies
		10:00 AM		"		
		10:30 AM		Coffee Break		
		11:00 AM		Ontologies: Presentation and Morphster	Morris/Sequeda	
		12:00 PM		Lunch		
		1:00 PM		Pack and load bus for travel to Guanacaste		
		1:30 PM		Bus to Hotel Boyeros		Hotel Boyeros, Liberia, Tel: 506 2666-0722
		3:00 PM		"		hboyeros@racsa.co.cr
		4:00 PM		"		http://www.hotelboyeros.com
		5:00 PM		"		
		6:30 PM		Dinner at Aroma Tico, on the way to Liberia		
		7:00 PM		Arrive Hotel Boyeros, Check-in		
		8:00 PM		**Session 19: Free Time in Liberia**		

	Sunday, June 08, 2008	7:00 AM		Breakfast				
	Hotel Boyeros, Liberia	8:00 AM		Bus to ACG				
		9:00 AM		**Session 20 (Data)** Area de Conservation Guanacaste (ACG), big picture of biodiversity		Dan Janzen/Winnie Hallwachs and 29 collaborators		
		10:00 AM		Biodiversity inventory: Up-close and personal, species, data and everything else		"		
		11:00 AM		"		"		
		12:00 PM		Lunch at ACG, Park Headquarters		"		
		1:00 PM		**Session 21 (Community Computing)** Sliding into DNA barcoding		"		
		2:30 PM		Coffee Break		"		
		3:00 PM		Taking the barcode revolution to an entirely new fauna		"		
		4:00 PM		"		"		
		5:00 PM		Free time (tentative)		"		
		6:00 PM		Bus back to Hotel Boyeros				
		7:00 PM		Dinner in Liberia				Recommended: La Toscana, Paseo Real (make reservation). Ronald will come back with recommendations
								la Toscana is in centro commercial across from BK
	Monday, June 09, 2008	7:00 AM		Breakfast, Check-out of Hotel				
	La Selva Biological Station	8:00 AM		Bus to La Selva				Some instructors depart Liberia Airport
		9:00 AM		"				
		10:00 AM		"				
		11:00 AM		Arrive La Selva				
		12:00 PM		Lunch				
		1:00 PM		**Session 22 (Tools) Working with Core Applications**		Morris		
		2:00 PM		Hands-on: Install and configure, PHP, MySQL		Morris/group		
		2:30		Coffee Break				

		PM						
		3:00 PM		Hands-on: Install Apache and Menalto gallery		Morris/group		
		4:00 PM		"		"		
		5:00 PM		Free time				
		6:00 PM		Dinner				
		7:00 PM		**Session 23 (Tools)** Hands-on: TAPIR install		Manuel Vargas		
		8:00 PM		"		"		
	Tuesday, June 10, 2008	6:00 AM		Breakfast				
	La Selva Biological Station	6:30 AM		Optional Guided Tour of Forest				
				Session 24 (Community Computing): Data				Impact of community data centers on science workflows
		8:00 AM		CRIA Brazil Community Computing		Dora Canhos		
		9:15 AM		Google Community of Applications and Services		Rebecca Shapely		
		10:00 AM		Coffee Break				
				TDWG Community: Standards and tools		Morris and Heidorn		
		10:30 AM		Google Community of Applications and Services		Rebecca Shapely		
		11:00 AM		"		"		
		12:00 PM		Lunch				
		1:00 PM		**Session 25 (Community Computing)** Encyclopedia of Life		David Shorthouse		EOL science workflows, architecture, tools, etc.
		2:00 PM		Global Collaborations through TAPIR		Manuel Vargas		
		2:30 PM		Coffee Break				
		3:00 PM		Global Collaborations through TAPIR (cont.)		Manuel Vargas		
		4:00 PM		GBIF Community		Mata		
		5:00 PM		Free time				
		6:00 PM		Dinner				

		7:00 PM		**Session 26 (Community Computing)** Content Management Systems - Drupal as an example		Shorthouse		
		8:00 PM		"				
	Wednesday, June 11, 2008	6:00 AM		Breakfast				
	La Selva Biological Station	6:30 AM		Optional Guided Tour of Forest				
				Session 27 (Community Computing):				
		8:00 AM		Open Source Software and Open Publishing		Morris		Open source licensing and creative commons licensing
		9:00 AM		Specify Software Project Network Architecture		Beach		
		10:00 AM		Coffee Break				
		10:30 AM		Software Design and Usability		Shapely		
		11:30 AM		OPEN				
		12:00 PM		Lunch				
		1:00 PM		**Session 28 (Data and Tools) Hands-on: Character-based identification**		Orlando Vargas/Enrique Castro		using online multiple-access keys for diagnostic IDs
		2:00 PM		Character-based Identification Concepts		Morris/Heidorn		Concepts, algorithms, technologies, tools
		2:30 PM		Coffee Break				
		3:00 PM		Continuation character data concepts and applications		Morris/Heidorn		Concepts, algorithms, technologies, tools
		4:30 PM		OPEN				
		5:00 PM		Free time				
		6:00 PM		Dinner BarBeQue at La Selva				
		7:00 PM		**Session 29 (Tools)** Spatial Analysis in Macroecology, SAM		Thiago Rangel		
		8:00 PM						
	Thursday, June 12, 2008	6:00 AM		Breakfast				

	Hotel Irazu	8:00 AM		**Session 30 (Various)** Course Wrap-up, Missed topics, Gaps Synergistic and Future Collaborations, and Next Steps	All		
	Best Western			PASI Impact Survey			
				Funding Opportunities: NSF and JRS foundations	Bryan Heidorn		
		10:00 AM		Coffee Break and Group Picture			
		10:30 AM		Future Collaborations (continued)			
		12:00 PM		Lunch and check out			
		1:00 PM		Check out			
		2:00 PM		Depart by bus for San Jose			
		3:00 PM		Bus			
		4:00 PM		"			
		5:00 PM		"			
		6:00 PM		Arrive San Jose and check in to Hotel Irazu			
		7:15 PM		Start boarding bus to restaurant			http://www.monastere-restaurant.com/
		7:30 PM		Bus to Le Monastere, Escazu			Santa Ana Road from Escazu, left at Multicentro Paco
							follow the green crosses.
	Friday, June 13, 2008	AM		International Departures SJO Airport			

7. References

Black C., et al. 1989. Loss of Biological Diversity: A Global Crisis Requiring International Solutions. National Science Foundation, National Science Board, NSB 89-171.

Environment Australia, 1998. The Darwin Declaration. Australian Biological Resources Study, Department of the Environment, Environment Australia, Canberra.

Foster, I. 2005. Service-Oriented Science. Science. 308: 814-817.

Lane, M. A. 1998. Teaming with Life: Investing in science to understand and use America's living capital. President's Committee of Advisors On Science and Technology: Panel on Biodiversity and Ecosystems. 86 pp.

NSF/OISE #0549443, Beach, J. H., and Ibarra, J. E. (2006). Improving International Biodiversity Research Collaboration with Cyberinfrastructure: A Report of the Workshop, "Cyberinfrastructure for International Biodiversity Research Collaboration", National Science Foundation., http://www.ciara.fiu.edu/biocyber/workshop_report.pdf

Sensor Arrays for Rainforest Research: Deploying a Wireless Sensor Network at OTS's La Selva Biological Station, Costa Rica, http://tracsvn.nhm.ku.edu/projects/otsmri/

Systematics Agenda 2000 (Committee). 1994. Charting the Biosphere. American Museum of Natural History, New York. 1994. 25pp. (Booklet)