**Biodiversity Cyberinfrastructure (BIOCI) Workshop**
**A Special Emphasis Component of the Southeast Asia Institute Program 2011**
**(SEAIP)**
**Taiwan, 3-4 December, 2011**

## Executive Summary

Biodiversity is the variation of life, ranging in scale from molecules, genes, cells, individual organisms, to species through ecosystems. Our understanding of the complexity and interaction of the planet's biodiversity is fundamental to a healthy and sustainable planet. The loss of biodiversity is a leading environmental and social issue as the human population grows, landscapes are modified, and our regional and global climate changes.

The Pacific Rim, and Southeast Asia in particular, are areas of impressive biodiversity, and also among the least well understood and studied. Southeast Asia is a region where there is high human and economic growth and therefore increased pressure on natural systems. Wallace's Line traverses the region, the multitude of islands such as those that make up the Indonesian archipelago, and the isolation effects of vast distances, pose intriguing and compelling science challenges. These are framed by the human cultural diversity of the region as well as networking, data sharing, and collaborative challenges.

More than ever before, the capacity for data production is increasing and the need to manage it is becoming problematic. From entire genomes being sequenced in a few hours to worldwide satellite images being produced on a daily basis, the capacity for institutions to host, analyze, standardize, and share their data is becoming the biggest challenge for biologists and their host-institution. For example, the timber industry is increasingly relying on the up-and-coming DNA fingerprinting technology. Soon, every tree felled will have its own DNA fingerprint to match along the custody chain. This type of data needs to be readily available to numerous agencies and laboratories for cross verification and certification. Throughout Southeast Asia, scientists are listing organisms in different parks and regions, recording invasion of sites by invasive species, restoring native species to reclaimed agricultural land, measuring the impact of anthropogenic activity on native ecosystems, among the many other areas of biodiversity research. The computational capacity, cost of maintenance and ability to share data with colleagues is nowadays a big part of any research budget. Especially in places that may not have the means of big research institutions, the need to coordinate efforts, agree on standards *a priori* and share facilities and budgets could make a crucial difference in scientific progress and societal benefit now and in the future. Currently, existing data sets are isolated within institutions, lack common standards for interchange, and this may mean that there are important biological questions or pressing environmental issues that could already be addressed, but technological and coordination barriers remain significant challenges. A structured and readily available cyberinfrastructure could tremendously enhance their research, and the impact of their results on education, management and policy.

The vision and goal that frame this workshop are to improve our understanding of the relationship of biological diversity to global change and sustainability science, with a focus on Southeast Asia, through building a network of researchers, policy makers, and citizen. To achieve this will require developing science-based educational, social, and environmental policies and programs for protection and sustainable use of natural resources through a

combined approach to research on biodiversity and cyberinfrastructure. In addition it will require a framework and set of tools for sharing experience, expertise, and data as well as building collaborations. A core challenge to realize these outcomes is how we address insufficient and inaccessible data and inadequate policies needed by scientists, decision-makers, and other stakeholders. We need to enable new, interdisciplinary approaches to research and problem-solving.

A collaborative approach in Southeast Asia, in biodiversity, and on cyberinfrastructure, is timely. First, several smaller efforts are starting (e.g. linking herbaria and collections digitization efforts in Indonesia and Malaysia; creating interdisciplinary teams to study the biodiversity of natural compounds in Malaysia; disseminating a Southeast Asia Biodiversity Informatics concept; and starting an AP-BON effort). Second, many economies in Southeast Asia are becoming more open to and supportive of research efforts. Third, with the increased network connectivity, the virtualization of resources for computing and data, access to technology, advancement of methods to link data (required for such a network) are more available to more researchers.

Both the cultural and professional diversity were represented in the participants of the workshop (there were researchers from Taiwan, Malaysia, Indonesia, Thailand, Philippines, Vietnam, in additional the United States, France and Germany; see Appendix 2). A set of research scenarios representing a broad range of interest areas were developed that could frame a broad vision for predicting biodiversity and climate change effects. These all emphasized a need to utilize, improve, and develop cyberinfrastucture that could lead to better research, collaboration, training, and access to data in the region. Areas of interest included natural products chemistry, linking data from heterogeneous resources, and building geospatial data capacity that could aid biodiversity and climate research.

Participants saw clear opportunities for new information technology approaches to study and integrate biodiversity data which comprise an increasingly important component of the data deluge. They saw a number of research and downstream benefits from gaining access to genetic, genomic data, remote sensing imagery, such as that from Formosa 2 and 3.

Participants agreed to continue to explore the research, educational, and policy issues of joint interest (noted above). The approach used focuses on a set of meetings over the next 12 to 18 months, with tangible outcomes from defined small projects between meetings. While there was a great deal of diversity of participants geographically throughout Southeast Asia, other future participants and partners were identified and will be invited to future workshops.

## Acknowledgements

Science Foundation (NSF OCI-0627026, PRAGMA).

## Workshop Outcomes

The workshop addressed overarching goals of improving our understanding of the relationship of biological diversity to global change and sustainability science in Southeast Asia. Ultimately, a network of researchers, policy makers, and citizens could achieve development of science-based educational, social, and environmental policies and programs for protection and sustainable use of natural resources through an approach that combines research on biodiversity and cyberinfrastructure. This will require a framework and set of tools for sharing experience, expertise, and data as well as building collaborations. Core challenges, to advance the needs scientists, decision-makers, and other stakeholders, were identified, that included addressing insufficient and inaccessible data and inadequate policies. We need to enable new, interdisciplinary approaches to research and problem-solving.

We explored biodiversity science challenges that go beyond traditional approaches to disciplinary biological research and natural resource management. We initiated new collaborations among researchers in Southeast Asia for innovative approaches to interdisciplinary, complex, data- and computationally-intensive biodiversity challenges. New scientific approaches require high-throughput data capture (e.g., sequences), distributed network integration of data capture devises (e.g., sensors), archives (museum collections and remote sensed imager), and compute resource, and advanced performance computing for analysis and visualization, archiving and dissemination.

The material in this short overview documents some of the key findings, presentations, and discussions at the workshop as well as some potential future activities. While not possible to capture in these note, there was a great deal of excitement by all the participants, who will look for ways to continue the discussions and implement some of the possible actions.

### Overview Day 1

There were four sessions of individual presentations on: Biodiversity informatics and data management, Chemistry of plant natural products, Forestry and other systems study, and Informatics (visualization, simulations, automation systems). See Appendix 1 for a list of presenters with topics, and a link to the presentations.

The researchers (see Appendix 2) presented valuable datasets on very diverse topics, such as mangrove systems, amphibians, entomology, medicinal compounds in herbs and woody plants throughout Sundaland[1] (including the Malay Peninsula on the Asian mainland, as well as the large islands of Borneo, Java, and Sumatra and their surrounding islands), validated drugs, butterflies overwintering habitat, invasive alien species, and permanent-geographically dispersed forest plots.

---

[1] The eastern boundary of Sundaland is the Wallace's Line, identified by Alfred Russel Wallace as the eastern boundary of the Asia's land mammal fauna. See http://en.wikipedia.org/wiki/Sundaland.

*Challenges*

During the presentation several challenges were articulated by the speakers, which are summarized here:

- Collections
    - How can the community or institutions manage collections in the long-term?
    - There are increasing numbers of large collections (GIS, genetic, etc…). How can an institution handle the storage of large data, and how can those dataset be shared with other?
- Sharing, Trust, and Collaborations
    - How do cultural barriers impede data sharing and building trust needed in successful collaborations?
    - What are some best practices for collaborations and establishment of trust (joint publications, property rights agreements, sharing facilities, funding, establishing long-term student programs)?
- Data managed with in-house software may be inaccessible and the tools may not scale well.
- Community Activities
    - How do we organize volunteers and staff for monitoring and standardizing data collections?
- Training
    - How can the community make biodiversity studies and research attractive for students? In some cases joint efforts between biology and social sciences that emphasize impact on society (poverty alleviation) can gain attention of some students.

## Overview Day 2

The Day 2 participants (Appendix 3) summarized challenges and split into three working groups. Although these summaries are incomplete, we are providing them to document some of the discussions and to suggest some possible activities for the broader community.

Each working group was asked to identify projects that would be of broad interest and which would achievable (at least if focused on smaller geographic regions) in a couple of years.

## Group1: Natural products chemistry
*Tangible Collaborative Projects*
1. Create a map of natural products (starting with a regional focus) in concordance with biodiversity researchers, to understand the diversity of natural products from a spatial perspective, and to compare with other geospatial measures of biodiversity.
2. Create metabolites profile and relate to traditional use via geospatial approaches

*Education & Outreach Projects*
1. Develop and conduct training materials and workshops in iNADI and NADI (Natural Products Discovery System[2]) for researchers and students.

---

[2] http://phds.usm.my/Backup/NADIdiscovery/index.php

2. Develop training on use of databases; and develop databases for others to use.
3. Develop programs to exchange of researchers & students between regional Institutes/Universities

*Stakeholders to invite in the future (partial list)*
1. Malaysian Society of Natural Product Chemistry
2. Indonesian Society of natural Product Chemistry
3. The Philippines, Thailand & Vietnam, Laos, Myanmar Societies of National product Chemistry

Participants: Shaharum Shamsuddin (lead), Habibah Wahab (Skype), Unang Supratman, Peter Arzberger

## Group2: Spatial data and analyses
### *Tangible Collaborative Projects*
1. Collaborations could be defined based on research questions that harness data infrastructure and opportunities with Taiwan space agency, e.g., access to Formosat 3 and possibly Formosat 2 (with funding). Access to such data could assist with
   a. Broad scale biodiversity modeling
   b. Microscale modeling (e.g., gaps in geospatial data for landscape level and below, e.g., community GIS) linked to socio-economic changes.
   c. Individual researchers who often find it difficult to access large scale geospatial data for various reasons.

2. There is still a need to identify existing research projects and people involved that would be interested in such work, including:
   a. Coral reef research
   b. Genomic Observatories initiatives
   c. Geospatial aspects of natural products chemistry
   d. Effects of climate change on ultramafic/serpentine vegetation along Pacific Rim, which as extreme environments may be highly susceptible to climate change.
   e. See list below for additional ideas discussed.

3. Write a white paper on methodologies and approaches related to socio-ecological integration

### *Infrastructure*
Harnessing cyberinfrastructure will be critical to make advances in this area. In this section the group discussed challenges and opportunities for CI, along with a type of grand challenge problem in this area.
1. Data warehousing challenges. The group considered that a linked data approach such as that described below would be useful for access to data and metadata in repositories.
   a. Remote sensing data from Formosat 2 and 3 could be very useful for biodiversity research in Southeast Asia. Sensors on these

5

vehicles, especially Formosat 3 are appropriate for climate research. The latter could be available freely for research.

    b. Taiwan has ongoing discussions with Thailand to provide real-time spatial information (e.g., for disaster recovery).

    c. Developing infrastructure to more easily share imagery for Southeast Asia research would be useful. The group considered uses by collaborations such as GLEON (the Global Lake Ecological Observatory Network, gleon.org), other ecological research and large-scale biodiversity research as good topics.

2. The data warehousing challenges are also opportunities. Ready paths forward were in

    a. Linked data, because others in the workshop were considering generic approaches, and while no solutions were yet in hand, spatial data provides a good use case.

    b. Biodiversity modeling and visualization are currently hot research topics, including species distribution, niche, and phylogeographic analysis

3. Grand Challenge problem

    a. Predicting the impact of climate change on biodiversity is both a scientific and social challenge.

Participants: Grace Wong (lead); Reed Beaman; Fang-Pang Lin; Parikesit

## Group3: Linked data: tools for sharing and standardizing data

The need to share data between institutions and research groups is greater than ever, and will only increase in the future. To be able to find, retrieve, and share data, the data needs to be linked. One of the most promising and implementable way of doing so is through the semantic web Standard RDF (Resource Description Framework, e.g., http://www.w3.org/RDF/). Each institution or laboratory can host their own data, but create query nodes. The standardization and agreement on ontologies (vocabulary) is essential. Biodiversity Information Standards (TDWG, http://www.tdwg.org/) has started this effort to implement a taxonomic data working group, but other complementary efforts are necessary. This needs to be implemented rapidly, ideally prior to significant local investments are made, since switching from an already well-established system can be complicated.

### Tangible Collaborative Projects

1. Create a working group to start model the data and set standards for ontologies.

2. Create a Southeast Asian biodiversity portal. You can have as many query portals as you want. Nobody owns it.

### Next steps

1. Launch several collaborations for pilot dataset. Several collaborations were agreed on during the session.

2. Experts workshop on Linked Open Data (LOD) in Southeast Asia: 1 week, in about 6 months-time, with different components:

    a. Training

b. Working group – ontology standards
c. Portal setup
3. Training in Dec 2012 in Taiwan. Aimed at less expert groups.

Participants: Cam Webb (Lead); Hsiu-Mei Chou; Charlotte Germain; Darlina Md Naim; Muchtaridi; Adam Murphy; Shahir Shamsir; Yu-Huang Wang


## Other possible partners – for engagement in future activities

This workshop was intended to explore whether there might be interest to continue regional discussions in biodiversity or aspects of biodiversity (e.g. Linked Data). Participants created a list of potential partners for future discussions.  We list these without intending any prioritization or appropriateness; furthermore none of the groups have been contacted. We would welcome interest from these or other groups.

- CBD Convention for Biological Diversity. 2012: Rio+10.
- State Biodiversity agencies
- National planning organizations
- Indonesian Society of natural products
- Double Helix
- International NGOs WCS, WWF, CI, IUCN
- AP-BON - GEOBON
- ACB Asian Center for Biodiversity (Philippines)
- Indonesia Science Institutes
- Industry and agroforestry
- RSPO (Roundtable for Sustainable Palm Oil) have HCVF assessments
- GBIF – nodes in SEA?
- Ministry of science Malaysia(MOSTI)
- Department of Fisheries resources Malaysia
- Forest Research Institute of Malaysia (FRIM)
- World Fish Center
- Participant organizations for education
- AP Bionet Pacific Rim South America – Based in Singapore – Bioinformatics, with one component of biodiversity
- APEC TEL- one working group on natural resources
- APAN
- Diversitas
- CBD Clearing House
- Ministry of Environment Indonesia
- BOLD Barcode of Life initiative
- GSC Genomics Sensus Consortium
- TDWG (Biodiversity Information Standards)
- Open Geospatial Consortium (OGC)
- Museums and herbaria regionally and globally

As a disclaimer we have not been in contact with any of these organizations.


## Conclusions and plans for the future

Both the cultural and professional diversity were represented in the participants of the workshop (there were researchers from Taiwan, Malaysia, Indonesia, Thailand, Philippines, Vietnam, in additional the United States, France and Germany). A set of research scenarios representing a broad range of interest areas were developed that could frame a broad vision for predicting biodiversity and climate change effects. These all emphasized a need to utilize, improve, and develop cyberinfrastucture that could lead to better research, collaboration, training, and access to data in the region.

Participants agreed to continue to explore the research, educational, and policy issues of joint interest (noted above). The approach used focuses on a set of meetings over the next 12 to 18 months, with tangible outcomes from defined small projects occurring between meetings. While there was a great deal of diversity of participants geographically throughout Southeast Asia, other future participants and partners were identified and will be invited to future workshops.

It is our sincere hope that this and other efforts will bring focus to the opportunities of biodiversity research in Southeast Asia and to the challenges inherent in general at the interface between biodiversity and cyberinfrastructure. We believe that through a combination of repeated discussions, concrete actions, and engagement of a broad, interdisciplinary community in Southeast Asia, that we, collectively, as a network, will be able to address major issues in biodiversity, train a new generation of researchers, and ultimately lend efforts to science-based policy.


## A Note of Thanks from the Organizers

As organizers of the workshop and recorders of the discussions, we have tried to capture accurately the key components of the discussions. Inevitably there are mistakes, which we assume responsibility for. We wish to extend our thanks to the funding agencies (National Science Council) for the bulk of the funding and the National Science Foundation through the PRAGMA award (NSF OCI 0627026) for their support, and to the National Center for High-performance Computing for organizing a wonderful venue. Finally we wish to thank the participants of the workshop for creating an environment to allow full discussions.

Should a reader of this report find errors, or want more information or want to participate, please contact us.

> Fang-Pang Lin, National Center of High – performance Computing,
>     fplin@nchc.narl.org.tw
> Reed Beaman, Florida National History Museum, University of Florida
>     rbeaman@flmnh.ufl.edu
> Charlotte Germain, Florida National History Museum, University of Florida
>     cgermain@ufl.edu
> Habibah Wahab, School of Pharmacy, Universiti Sains Malaysia
>     bibwahab@gmail.com
> Peter Arzberger, University of California San Diego parzberg@ucsd.edu


## Postscript (21 May 2012)

Since this workshop several additional steps have been undertaken that were motivated by discussions described above. In particular, several of the members of the Spatial Data and Analysis Group put together a one day workshop, Biodiversity in (Southeast) Asia Workshop: Defining Cyberinfrastructure to Support the Study of Ultramafic Ecosystems, that was held in conjunction with PRAGMA 22 workshop, on 17 April 2012, in Melbourne Australia; host Monash University (http://pragma22.pragma-grid.net/dct/page/70005). This workshop was a specific explorative activity focused on biodiversity in ultramafic regions. The workshop produced activities (data provenance, and some preliminary analysis with Lifemapper) to be completed by PRAGMA 23 (10 – 11 October 2012).

## APPENDICES

### Appendix 1: List of Presentations on Day 1: Saturday 3 December 2011

See http://event.nchc.org.tw/2011/southeast_asia/index.php?CONTENT_ID=2 to find versions of the presentations.

| Name | Institution | Title of Presentation |
|---|---|---|
| Overview | | |
| Arzberger, Peter Beaman, Reed Lin, Fang-Pang | University California San Diego University of Florida National Center for High-performance Computing | Overview of Biodiversity and Cyberinfrastructure |
| Biodiversity Informatics | | |
| Beaman, Reed | Florida Museum of Natural History, University of Florida | Biodiversity science: A new, unique cyberinfrastructure challenge: Or familiar, generic problem space? |
| Shamsir, Shahir | Bioinformatics Research Group, Biological Sciences Department, Universiti Teknologi Malayisa | Overview of Biodiversity Activities at UTM |
| Lu, Sheng Shan Wang, Yu-Huang, Lin, Chau-Chin | Ecological Informatics Research Group, Taiwan Forest Research Institute | Taiwan Experience in building, managing and applying ecological and biodiversity information systems |
| Chemistry and Natural Products | | |

| | | |
|---|---|---|
| Shamsuddin, Shaharum Wahab, Habibah | Advance Molecular and Cellular Biology Laboratory Universiti Sains Malaysia School of Pharmaceutical Sciences, USM | Leveraging Malaysia's Biodiversity towards Value Creations using Bioinformatics |
| Supratman, Unang | Department of Chemistry, Universitas Padjadjaran | Biologically Active Natural Products from Indonesian Plants |
| Lin, Jung-Hsin | Institute of Biomedical Sciences, Academia Sinica; School of Pharmacy, National Taiwan University | Exploring the Biodiversity in the Traditional Medicine: an in silico Target Identification Approach |
| Muchtaridi | Universitas Padjadjaran | Pharmacoinformatic for Drug Discovery from Natural Product |
| Biodiversity and Forestry | | |
| King, Hen-biau | Taiwan Ecological Research Network | Old Questions, New Problems, Challenges & Opportunities |
| Chao, Jung Tai | Taiwan Forest Research Institute | On Sharing |
| Webb, Cham | Harvard University | Informatics Interests |
| Germain, Charlotte | University of Florida | Phylogeographer and Conservation Geneticist |
| Murphy, Adam | Global Diversity Foundation | Ongoing GDF work in Sabah |
| Md Naim, Darlina | Universiti Sains Malaysia | Centre for Marine and Coastal Studies (CEMACS) |
| Pampang Parikesit | Institute of Ecology, Universitas Padjadjaran | Biodiversity Conditions and Ecosystems Services across gradient of human-modified landscape in Agricultural Landscape of Humid Tropic West Java Indonesia |
| Chumkiew, Sirilak | NECTEC | Center of Excellence for Ecoinformatics NECTEC-Walailak |
| Wong, Grace | United Nations Development Programme (UNDP) | Building Sustainable Landscapes |
| General Informatics | | |
| Lin, Fang-Pang | NCHC | Informatics Interests |

| | | |
|---|---|---|
| Johnson, Andy | University of Illinois at Chicago | Research Related to Biodiversity |
| Woessner, Uwe | HLRS | Current Projects, Challenges, Opportunities |
| Chayopitak, Nattopon | EGAT-NECTEC | EGAT and Biodiversity in Reservoirs |

## Appendix 2:  Participants on Day 1: Saturday 3 December 2011

| Name | Organization | Location |
|---|---|---|
| Boris Arnoux | Télécom-ParisTech | France |
| Peter Arzberger | UCSD | US |
| Nuallaor Banomyong | EGAT | Thailand |
| Reed Beaman | FLMNH | US |
| Decha Boonyakanjana | EGAT | Thailand |
| Sam Chanrathany | Hanoi University of Technology | Vietnam |
| Jung Tai Chao | Taiwan Forest Research Institute | Taiwan |
| Nattapon Chayopitak | NECTEC | Thailand |
| Hsiu-Mei Chou | NCHC | Taiwan |
| Sirilak Chumkiew | Wilailak University | Thailand |
| Do Duc Dong | Vietnam National University | Vietnam |
| Nena Carina Espanola | ASTI | Philipine |
| Charlotte Germain | FLMNH | US |
| Andy Johnson | EVL | US |
| Seubsuang Kachapornkul | NECTEC | Thailand |
| Henbiau King | Taiwan Ecological Research Network | Taiwan |
| Ling-Ling Lee | National Taiwan University | Taiwan |
| Jung-Hsin Lin | Academia Sinica | Taiwan |
| Fang-Pang Lin | NCHC | Taiwan |
| Sheng-Shan Lu | Taiwan Forest Research Institute | Taiwan |
| Darlina Md Naim | Universiti Sains Malaysia | Malaysia |
| Muchtaridi | Universitas Padjajaran | Indonesia |
| Adam Murphy | Global Diversity Foundatoin | Malaysia |
| Nguyen Thuy Le | Ministry of Education and Training | Vietnam |
| Parikesit | Universitas Padjajaran | Indonesia |
| Pham Hong Phong | Hanoi university of Engineering and Technology | Vietnam |
| Pasit Sagganayok | EGAT | Thailand |
| Shahir Shamsir | Universiti Teknologi Malaysia | Malaysia |
| Shaharum Shamsuddin | Universiti Sains Malaysia | Malaysia |
| Unang Supratman | Universitas Padjajaran | Indonesia |
| Sirichet Suttiwong | EGAT | Thailand |
| Jelina Tetangco | ASTI | Philipine |
| Kanokvate Tungpimolrut | NECTEC | Thailand |
| Habibah Wahab | USM | Malaysia |

| Yu-Huang Wang | Taiwan Forest Research Institute | Taiwan |
| Cam Webb | Harvard | US |
| Uwe Woessner | HLRS | Germany |
| Grace Wong | UNDP Lao PDR CO / Ministry of Planning and Investment | Lao |

## Appendix 3:  Participants on Day 2: Sunday 4 December 2011

| Name | Organization |
| --- | --- |
| Peter Arzberger | UCSD |
| Reed Beaman | FLMNH |
| Hsiu-Mei Chou | NCHC |
| Charlotte Germain | FLMNH |
| Fang-Pang Lin | NCHC |
| Darlina Md Naim | Universiti Sains Malaysia |
| Muchtaridi | Universitas Padjajaran |
| Adam Murphy | Global Diversity Foundation |
| Parikesit | Universitas Padjajaran |
| Shahir Shamsir | Universiti Teknologi Malaysia |
| Shaharum Shamsuddin | Universiti Sains Malaysia |
| Unang Supratman | Universitas Padjajaran |
| Habibah Wahab | USM |
| Yu-Huang Wang | Taiwan Forest Research Institute |
| Cam Webb | Harvard |
| Grace Wong | UNDP Lao PDR CO / Ministry of Planning and Investment |