STRATEGIES FOR OPEN AND PERMANENT ACCESS TO SCIENTIFIC INFORMATION IN LATIN AMERICA: FOCUS ON HEALTH AND ENVIRONMENTAL INFORMATION FOR SUSTAINABLE DEVELOPMENT

ATIBAIA, SÃO PAULO, BRAZIL

MAY 08 – 10, 2007

Dora Ann Lange Canhos, Vanderlei Perez Canhos, William Anderson, Paul F. Uhlir, Bonnie C. Carroll, Raed M. Sharif, and Puneet Kishor, Editors

ORGANIZERS:

Centro de Referência em Informação Ambiental (Reference Center on Environmental Information - CRIA)

U.S. National Committee for CODATA

Brazilian National Committee for CODATA.

CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries

SPONSORS:

Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq

Brazilian Academy of Sciences (ABC)

U.S. National Science Foundation

Brazilian National Institute for Space Research (INPE)

Centro de Referência em Informação Ambiental (Reference Center on Environmental Information - CRIA)

U.S. National Weather Service

The Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries, of the United Nations Global Alliance for ICT and Development, UN G@ID

Inter-American Biodiversity Information Network

WORKSHOP REPORT

CONTENTS

The table of contents is empty because none of the paragraph styles selected in the Document Inspector are used in the document.

INTRODUCTION

This workshop provided an international and interdisciplinary forum to promote discussions and a deeper understanding of the requirements for long-term preservation, open access, and improved use of digital scientific information resources for sustainable development in Latin America. The workshop focused on health and environmental information for sustainable development and had four primary objectives:

i. Review and discuss the current status of practices for providing open availability of and permanent access to scientific information resources;

ii. Identify and discuss scientific, legal and policy, institutional and economic, and management and technical factors relevant to permanent access to digital scientific information resources;

iii. Identify follow-up activities that can be taken towards improving permanent access to the major type of digital scientific information resources; and

iv. Provide a networking opportunity for future collaborative activities for workshop speakers and attendees across discipline, institutional, and national boundaries.

The workshop was divided into three sessions: plenary informative sessions, thematic breakout sessions, and crosscutting breakout sessions.

There were 75 participants: 37 from Brazil, 14 from USA, 6 from China, 3 from Colombia, 2 from Chile, and one from each of the following countries: Cuba, Jamaica, Venezuela, Peru, Ecuador, France (French Guyana), Canada, Belgium, Denmark, UK, India, South Africa, Japan.

PART I. PLENARY INFORMATIVE SESSIONS

The first part of the workshop provided background information on the main issues that were to be discussed. Speakers presented their views about current open access activities and prepared participants for the crosscutting breakout sessions that followed.

The keynote speaker was Dr. João Vianei Soares, Director for Earth Observations of Brazil's National Institute for Space Research (INPE) who addressed the *role of scientific information in sustainable development*, focusing on Latin America[1]. He stressed the benefits of the Brazilian space program to sustainable development and INPE's role in making data and information available to society for a variety of uses and applications such as fire alert, deforestation maps and weather and climate forecast, all freely and openly available on the Internet. INPE is also developing a geo-information software package and making it available under an open source license. Cooperative networks include health, public security, urban planning, and natural disasters. INPE also has a very important program on satellite imagery in collaboration with China (Chinese-Brazilian Earth Resources Satellite Program, CBERS) and both countries have agreed to offer data at no cost to Latin America and Africa.

The four sessions that followed were devoted to (i) open and permanent access to scientific information; (ii) policy and legal issues; (iii) institutional and economic issues; and (iv) management, technical and infrastructure issues.

Session 1. Open and Permanent Access to Scientific Information: current challenges and strategic directions

This session was chaired by John Wilbanks, Science Commons, USA. Panelists were:

---

[1] Available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session1/jvianei.ppt

- Carlos Ferreira de Morais Pires (head of sector for Scientific Data in the Information Society and Media Directorate General of the European Commission): *e-Infrastructures: the European Perspective on Scientific Data*[2];
- LIU Chuang (Chinese Academy of Sciences): *Progress of Open Access to Scientific Data in East Asian Countries*[3];
- Steve Rossouw (South African National Committee for CODATA): *African perspective of Open & Permanent Access to S&T Information*[4]; and
- Heather Joseph (Scholarly Publishing and Academic Resources Coalition - SPARC). *A Question of Access: Public Policy and Digital Repositories*[5].

Carlos Morais Pires stressed the importance of data, information, and knowledge to boost scientific excellence in Europe. He presented the steps taken at the European level to empower research using ICT based infrastructures for scientific information. Other points included scientific data as infrastructure, its use in and between different disciplines, and its dissemination through education.

LIU Chuang presented the progress of open access to scientific data in East Asian countries. She showed the change of attitude towards data in China. There were very few people dealing with data archives before the 1980s but a growing perception of data issues began and in 2000 the current digital "storm" was recognized and information infrastructure was expanded. Prof. Liu also described the "*UN Global Alliance for ICT and Development (UN GAID) Community of Expertise: Education Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries (e-SDDC)*" and its mission to bridge research, education, and policy on scientific data in developing countries under the UN GAID framework. The idea is to create a forum or platform for dialogue for all stakeholders to: identify and evaluate different mechanisms and policies for promoting greater access to and use of digital S&T data for meeting the needs of developing countries in policy reform in scientific data management and applications; and to help build a distributed and decentralized network of networks on scientific data and information resources for innovative research, sustainable development, and better life in the developing world. Longer term goals include: (i) bridging scientific research, education, and policy decision making in order to find tailored solutions for problems faced in developing countries; (ii) enhancing the use of ICT in the application of S&T information for scientific innovation and sustainable development; and (iii) applying these digital resources to the UN Millennium Development Goals in developing countries.

Steve Rossouw reviewed current strategies on permanent access to S&T information in Southern Africa, where a number of workshops took place. Recommendations include the need to: influence government officials; raise awareness as to data value; hold workshops on data handling; offer outreach activities such as regional workshops; and the need to include socio-economic data.

Heather Joseph's talk was on open access journal publishing, focusing on alternatives to the traditional "user pays" model of supporting peer-reviewed journals. Open access repositories, which focus on the creation of freely accessible digital databases, populated by the output of individual researchers and organized around disciplinary or geographical (institutional) constraints were presented as an alternative. An important point made is that government-wide policies, geared towards ensuring that the value of publicly-funded research is maximized, have begun to emerge worldwide. These policies are beginning to explore the possibilities for advancing the conduct of research by providing stable, long term archives for research outputs, and maximizing access and use of research results.

Session 2. Policy and Legal Issues

This session was chaired by Regina Gusmão (Centro de Gestão e Estudos Estratégicos – CGEE, Brazil). Panelists were:

---

[2] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session2/cmorais.ppt

[3] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session2/lchuang.ppt

[4] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session2/srossouw.ppt

[5] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session2/hjoseph.ppt

- Dora Ann Lange Canhos (CRIA): *Open access to data and information: the need for strategies and policies in Brazil[6];*
- Ana Maria Prat (CONICYT, Chile): *Towards a National Policy to Provide Access to Scientific Information and Data in a Developing Country[7];*
- Harlan Onsrud (Department of Spatial Information Science and Engineering, University of Maine, USA): *Legal Constraints and Opportunities in Providing Permanent Access to Scientific and Technical Data[8];* and,
- Pedro Paranaguá (Fundação Getúlio Vargas, Brazil): *Creative Commons: An alternative for sharing scientific knowledge[9].*

Dora Canhos focused her talk on species and specimen data and on the necessity of developing policies to guarantee open and free access and long-term archiving to non-sensitive scientific data. Both international and national scenarios were presented and the opportunities and the need to coordinate policies concerning data providers, custodians, and financing agencies were stressed. Recommendations included the necessity of having policies for data that is born digital, of having mechanisms for digitizing legacy data, and of implementing long-term policies for the development and maintenance of data centers.

Ana Maria Prat emphasized the need to define a legal framework to ensure both access to foreign scientific information and the diffusion and accessibility of information and data produced locally. She stressed the perspective of open access to a developing country, that does not have an editorial industry and that cannot afford access to the world's information. The approach to open access is one of extreme necessity, not only to be able to access the world's knowledge, but to make the country's research results available. Ana Maria concluded by saying that information is largely available but not always accessible, so an integrating program is needed to allow interoperability of all the information available. It is important to make use of new technologies to assure better access.

Harlan Onsrud stated that the ability to collect and preserve scientific and technical data is affected by intellectual property rights that others may have. One approach would be for creators of collections of scientific and technical data to always convey to the world any rights they may have in a dataset or database through a public domain or open access license (e.g., Creative Commons licenses). This is likely to require a global networked scientific infrastructure that provides benefits to those who make their data openly available.

Pedro Paranaguá spoke about Creative Commons licenses that are a way to foster diffusion and sharing of knowledge without giving away one's copyright. This is a more flexible way of protecting works without preventing follow-on innovation and creativity.

Session 3. Institutional and Economic Issues

This session was chaired by Holm Tiessen, Director of the Inter-American Institute for Global Change Research (IAI). Panelists were:

- Paul Uhlir (International Scientific and Technical Information Programs at the National Academies in Washington, D.C.): *Emerging Institutional Paradigms for Providing Open and Permanent Access to Scientific Data and Information[10].*
- Lewis Greene (Editor of the Brazilian Journal of Medical and Biological Research, and Professor at São Paulo State University): *Effect of Open Access on Brazilian Academic Journals[11].*

---

[6] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session3/dcanhos.ppt

[7] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session3/aprat.ppt

[8] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session3/honsrud.ppt

[9] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session3/pparanagua.ppt

[10] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session4/puhlir.ppt

[11] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session4/lgreene.ppt

- Barbara Kirsop (Electronic Publishing Trust, UK): *Institutional Repositories and their contribution to sustainable development[12]*.
- Imre Simon (Mathematics and Statistics Institute of São Paulo State University): *The Virtual Incubator of Digital Contents of the Tidia Program of FAPESP* (The State of São Paulo Research Foundation)[13].

Paul Uhlir argued that the pre-internet legacy business model for research journals and data centers that restricts access to and use of publicly funded scientific information does not take advantage of the potential benefits of global open availability online and consequently is slowing the progress of science and socioeconomic development. The question, therefore, is not whether open availability to such publications is better than access provided on terms that are economically, legally, and technologically restricted, but how effectively and quickly can open availability be institutionalized. He asserted that a gradual, but highly significant and far-reaching restructuring of scientific communication in public research is taking place. In his talk, Paul Uhlir stated that openness should be the default rule, subject only to legitimate and well-justified exceptions. This is increasingly the policy being adopted in many research sectors, institutions, and countries.

Lewis Greene emphasized the need for a journal to have visibility and availability. In developed countries, visibility and availability are linked by effective libraries supported by government grants for "overhead" to universities to pay for institutional subscriptions to journals. This situation does not exist in developing countries. An initiative in Brazil called SciELO (Scientific Electronic Library Online) provides open access to a number of journals and has increased the readership and citations of some scientific journals published in Brazil, but not to the same extent for all. An issue debated was the value of publishing in national versus international scientific journals. In order to increase scientific productivity indicators, the best Brazilian science is published in high impact international journals. But publishing in national scientific journals is fundamental to communicate research results of national or regional interest and to stimulate the development and consolidation of research fields within Brazil.

Barbara Kirsop stated that research progress is severely hampered by the cost of accessing scholarly journals containing essential information. The open access movement has generated a number of solutions, some of which are long-term, but the low cost and simple establishment of interoperable institutional repositories provides free access to published research and a mechanism for raising the visibility of national research findings. These repositories are supplementary access tools that provide increased impact to an organization's research output and to funder's investments. As the value of this strategy is recognized, institutions and funding organizations are mandating the archiving of published research in institutional or central repositories. For information-deprived scientists, this development offers opportunities to reach scientific independence, remove professional isolation, and contribute to the global knowledge pool.

Imre Simon presented the *Virtual Incubator project* that aims to promote and facilitate cooperative developments of open digital content of academic, technological, or social relevance. It provides, through the Internet, a widely available computational infrastructure that allows user communities to focus on quality content production instead of system maintenance. The Incubator is in operation since August, 2004 and in April 2007 had 400 projects and 11 thousand users. The system is compliant with the Open Archive Initiative and could act as an institutional or thematic repository.

Session 4. Management, Technical and Infrastructure Issues

This session was chaired by Michael Stanton, from Brazil's National Education and Research Network (RNP). Panelists were:

---

[12] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session4/bkirsop.ppt

[13] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_08-05/Session4/isimon.pdf

- Raymond McCord (Oak Ridge National Laboratory, USA): *Management Issues for Permanent Access to Scientific Information: Increases in Programmatic Scale and Change*[14].
- Martha Giraldo (RENATA, Colombia): *RedClara, an infrastructure for collaboration*[15].
- Marcio Faerman (RNP, Brazil): *Technology and Infrastructure Support for Large Scale Information*[16].

Raymond McCord stressed the fact that procedures for preparing information for permanent access should be merged with research planning as part of a "modern scientific method" and should receive similar incentives. Management should reinforce these practices by insisting on early planning for permanent access (data archiving) and providing specific rewards for these activities. Other management issues include protecting initial discovery opportunities, supporting long-term stewardship of data, and providing "cross training" of archive personnel in both scientific and information disciplines.

Martha Giraldo presented *RedClara* the Latin American network for Science, Education and Innovation. This initiative has established a nongovernmental organization (CLARA) that represents the interests of the related institutions and is developing an infrastructure that integrates the Latin American academic networks. *RedClara* is connected to the European Advanced Network GÉANT, to INTERNET2 (USA), and to academic networks in Latin America. CLARA promotes direct integration of scientific communities in diverse countries and foments the development of collaborative research and educational projects among them. CLARA's main objective is to integrate regional telecommunications networks and interconnect the National Academic Networks (NRENs) of the region in order to accelerate the development of the information society in Latin America.

Marcio Faerman spoke about the challenge of managing enormous volumes of data in a reliable and efficient manner. Emerging technologies integrating sensors and high end computing systems are generating a massive amount of data and the extraction of meaningful information from raw data files is complex and often requires multi-institutional as well as multi-disciplinary collaboration. Technology must deal with: tools for analysis (large data volumes can be generated much faster than they can be analyzed), scalability, data integrity (at terabyte scales, failures and data corruption are very likely to occur); efficiency; distributed access; and infrastructure management. Challenges include extracting knowledge from raw data, allowing cross-institutional and cross-disciplinary collaborations, and working with multiple formats. One technological response is an *information grid*, an integration of computers, communication, storage, and instrument resources into a powerful infrastructure. Infrastructure, in this case, is heterogeneous, distributed, and very complex. Middleware and data oriented tools may act as facilitators to tackle data management complexities. Integration between domain scientists (both data users and providers) is still a challenge that requires much more cross-disciplinary interaction. Marcio Faerman concluded indicating that information infrastructure is being redefined in Brazil and Latin America and that now is the time to have as much cross-disciplinary interaction as possible to define needs, partnerships and investments.

Technical Demonstration

The event opened an extra session for technical demonstrations. Antônio Mauro Saraiva and Pedro Luiz P. Corrêa both from the Polytechnic Institute of the State University of São Paulo respectively presented: WebBee - an information network on bee biodiversity in Brazil[17] and ViNCES - Weblabs on ecosystem services[18].

---

[14] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session6/rmccord.ppt

[15] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session6/mgiraldo.ppt

[16] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session6/mfaerman.ppt

[17] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session9/asaraiva.ppt

[18] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session9/pcorrea.ppt

Pollinators are critical species for maintaining ecosystems equilibrium and for agricultural production. Their decline has motivated many actions and initiatives, from local to global, to increase knowledge and to promote their conservation and sustainable use. WebBee is a network that involves researchers from many Brazilian institutions and provides information about pollinating species focusing on native Brazilian bees. It includes texts, images and videos about bee biology, beekeeping and management and relations with the plants visited. In order to increase this comprehensiveness, WebBee will be integrated to regional and global biodiversity information systems to exchange information through the Internet.

Web labs are real or virtual laboratories that are available through the Internet for remote access. They promote collaboration and sharing of experiments among researchers and allow the use of experimental facilities for distance education. The ViNCES project uses that technology for research and education on themes related to two ecosystem services: pollination and photosynthesis. Experiments about bee behavior using audio, video and environmental variables monitoring, and studies on photosynthesis under CO_2-enriched atmosphere in open-top chambers can be controlled or supervised remotely via web. The interconnection of research groups involved in the project through an advanced network allows overcoming barriers imposed by conventional networks. This Advanced Network is provided by the TIDIA-KyaTera[19] project, from FAPESP, in which the ViNCES project is inserted.

PART II. THEMATIC BREAKOUT SESSIONS

Thematic breakout sessions focused on data and metadata in the fields of health and biomedicine, earth and environmental science, biodiversity, and related journals. These sessions built upon the results of the preceding plenary discussions. Questions for speakers included:

- What are the current challenges and barriers to providing open and permanent access to data and information in this field?
- What are some of the most promising existing models or mechanisms for providing such access?
- Identify one or more potential realistic cooperative activities in Latin America for subsequent discussion at the breakout session.

Session 1. Health and Biomedical Data and Metadata

Anna Maria Prat (Conicyt, Chile) chaired this session and William Anderson (Praxis101, USA) was the rapporteur. The following presentations were given prior to discussions:

- Carlos Castillo-Salgado (Pan American Health Organization - PAHO): *Development and use of core public health indicators for decision making in the public health sector. The PAHO's experience[20]*.
- Adalberto Tardelli (Bireme, Brazil): *Scientific Literature for the Interagency Network of Health Information* - RIPSA[21],
- Laís Costa[22], (Fiocruz) and Janette Aguirre[23] (ACTO - Amazon Cooperation Treaty Organization): *PanAmazonian Network of Science, Technology and Innovation in Health*

---

[19] for more information see www.**kyatera**.fapesp.br

[20] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/ccastillo.ppt

[21] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/atardelli.ppt

[22] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/lcosta.ppt

[23] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/jaguirre.ppt

- Jane Costa (Fiocruz, Brazil): *Ecological Niche Modeling: A tool set to assess distributional patterns in biodiversity and pathogens*[24],
- Luiz Ary Messina (RNP/RUTE, Brazil): *Infrastructure perspectives for the development of Telehealth in Brazil and Latin America*[25],

Carlos Castillo-Salgado, from PAHO presented a set of core public health indicators developed in 1995 through a regional initiative for monitoring major health programs and mandates in the Americas. This initiative promotes the use of a selected set of highly standardized and validated health information. A total of 118 consensually derived indicators from demographic, social, economic, mortality and morbidity, human resources, access, and data on coverage of health services have been included in the database, the majority disaggregated by age and sex. These sets of indicators assist policy makers in the characterization of the health situation in 48 countries and are used for monitoring the situation of priority health problems.

Key challenges for open and shared access to health and biomedical data and metadata include:
- lack of core health indicators and infrastructure for handling system outputs and its distribution in the Americas - PAHO has built such an infrastructure, but this is required everywhere;
- lack of a sufficient number of trained people;
- limited access to critical information - most countries share data, but it can take years to get information;
- limited political support and financing; and
- the need for more use and analysis, which is also related to training and support.

Adalberto Tardelli presented the *Rede Interagencial de Informações para a Saúde* (RIPSA), an inter-institutional cooperation initiative established 10 years ago to support the use of information in the national health system in Brazil. The initiative assembled a set of basic indicators suitable for carrying out general analysis of the national health situation and its trends, with clear definitions of concepts and meanings, data sources and calculation methods, as well as their uses and limitations, and available and recommended analysis categories. Through a collaborative work process involving primary institutions, this set of basic indicators and data is annually reviewed, updated and made available in an easily accessible database over the Internet. RIPSA provides DeCS (Health Sciences Descriptors) a service for searching scientific literature for 101 basic indicators. The system applies an open literature-based discovery approach to find the concepts occurring in a cluster of articles. The initiative resulted in the development of a Virtual Health Library available at www.bvsalud.org.

Janette Aguirre provided background information on the Amazon region, which represents more then 40% of South America. This region is strategic in the realm of global geopolitics and has been of vital importance for local and global human health, due to the changes and transformations of its ecosystems. Added to the fact that the region includes eight countries and French Guyana, with an area of about 7.5 million square km and a population of about 30 million, it demands integrated public policies and simultaneous actions concerning infrastructure, social and economic development, and environmental management. The Amazon Cooperation Treaty Organization (ACTO) was established in 1978 by Bolivia, Brazil, Colombia, Ecuador, Guyana, Peru, Suriname and Venezuela, with the intention of promoting joint actions for the sustainable development of the Amazon basin. The treaty is a strategic political instrument for regional cooperation and plays a key role in confronting the common challenges of regional development and in improving the living standards of the population. It is a political forum with regional positions in global negotiations. Aiming at sustainable development of the Amazon region, strategic actions include: (i) conservation and sustainable use of renewable natural resources; (ii) knowledge management and technological exchange; (iii) regional integration and competitiveness; and (iv) institutional strengthening.

Challenges include:
- generating political will for integration and coordination of national and international organizations in the development of the Amazon (from the local perspective);

---

[24] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/jcosta.ppt

[25] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.1/lmessina.ppt

- promoting human development;
- building consensus on a health plan for the Amazon;
- promoting the work of regional networks; and
- increasing institutional capacity for handling and using health information in appropriate ways for political and technical decisions.

Laís Costa from the Fundação Oswaldo Cruz, Brazil, stressed the fact that building science, technology, and innovation in health research skills is a local, regional, and global necessity. Challenges must be addressed in collaboration with other Amazonian countries, and if needed with other countries, development agencies, and initiatives. The health sector has the potential to facilitate collaborative solutions. The Pan-Amazonian Network of Science, Technology & Innovation in Health (ST&I/H) is viewed as a way to achieve the goals of the Amazon agreement. It is supported by the Oswaldo Cruz Foundation (FIOCRUZ), the Amazon Cooperation Treaty Organization (ACTO), the Pan-American Health Organization (PAHO) and the Association of Amazonian Universities (UNAMAZ). Challenges for the region include:

- establishing the region's importance in a global context;
- implementing public policies to develop science, technology, biotechnology and innovation skills;
- fostering civil society's participation in defining priorities and evaluating policy implementations;
- reaching national and South American developments of the health sector through a strategic approach; and
- establishing natural resources management practices.

The Pan-Amazonian Network of ST&I/H intends to integrate institutions of education, research and management from the eight Amazonian countries – Bolivia, Brazil, Colombia, Ecuador, Guyana, Peru, Suriname and Venezuela – and debate the accomplishment of joint actions.

Aims of this network include: (i) improving science and technology capacity, to apply research to social needs; (ii) setting Amazon knowledge-based priorities; (iii) improving regional capacities; and (iv) supporting production, rationalization, dissemination, application, and evaluation of S&T in the region.

Jane Costa spoke about ecological niche modeling as a growing field with many potential applications in geography and ecology of disease transmission. This presentation reviewed relevant applications of ecological niche modeling in understanding the ecology of diseases; in characterizing areas of potential distribution; in identifying areas of potential invasion; in anticipating areas of risk due to climate change; in identifying unknown vectors or hosts; and in assessing data sets.

Luiz Ary Messina presented an ongoing project in Brazil, run by RNP the National Education and Research Network that is building municipal area networks in all state capitals connecting major public universities and research centers in the country. This infrastructure will be available at Rute[26], a Telemedicine University Network funded by the Ministry of Science and Technology. This will enable participating hospitals to use the network to run telemedicine and telehealth applications including video and web conferencing, second opinion, and continuous education. New strategies and perspectives are being created to integrate Latin American national programs and initiatives in telehealth to the USA and Europe.

Challenges in telemedicine practices include ethics, risks, responsibility, payment, licensing, qualification, standards, and education. Opportunities include the connection with the Mexican Telemedicine program and the Ecuadorian Advanced Academic Networks.

Discussion

After the presentations the group discussed a number of questions and topics as follows.

---

[26] for more information about RUTE (Rede Universitária de Telemedicina) see www.rute.rnp.br

1. **What are some of the most promising existing models or mechanisms for providing open access to health information, data and metadata?**

The group agreed that it is necessary to build networks to integrate and facilitate collaboration and contacts among researchers and institutions. Training courses are also important to enable professionals to use new technologies and practices related to providing open access. Workshops and events to discuss policies and rules regarding open and permanent access of data and information are also important. Another point was to try to make politicians aware of the difficulties that restrictive laws cause to scientific research and policy development, and to create financial support for the objectives of open access policy.

2. **What are some potential realistic cooperative activities in Latin America to help address the challenges and barriers that have been identified? The following specific activities were developed during this session.**

GEOSS (Global Earth Observation System of Systems) was identified as a partner for Ecosystems, climate and health projects. A demonstration project could involve linking PAHO public health indexes with climate indexes (e.g., Dengue fever and temperature).

Activities could include:
- Bibliographic database of Latin American papers on Dengue fever
- Monitoring disease vector (transmission) activity, using both legacy and new data types
- Digitization and dissemination of old health records
- Generating scenarios for interventions (local, national, regional)

The following institutions are interested in participating in such a project: ACTO, FIOCRUZ, PAHO, INPE, and NOAA. Laís Costa (FIOCRUZ), Janette, Aguirre (ACTO) would be interested in coordinating this activity.

Another possible project could be a ***university telehealth network*** in Latin America, based on national telehealth programs of Brazil and Mexico and on initiatives in Columbia, Ecuador, and Chile. This could involve the following organizations: *Red*Clara, RNP, CENTEC, CEDIA, FUNDATEL, RENATA, FIOCRUZ, ACTO, PAHO, IDB, and WB.

Practical steps include: (i) forming a working group on telehealth under *Red*Clara; (ii) networking institutions; and (iii) promoting video conference sessions. The idea would be for each Latin American country to submit a proposal for strengthening both national and local capacity. Coordinators could be Luiz Ary Messina (Instituto de Informática e Automação – IIA, Brazil) and Carlos Castillo-Salgado (Pan American Health Organization, USA).

The following activities could be included: (i) capacity building and training in health & biomedical data; (ii) internships for young professionals and sabbaticals for experienced professionals (international, interdisciplinary sabbaticals); (iii) link virtual campuses in Latin America; and (iv) list available on-line programs on health and environment

General recommendations were also made to national science academies. They include:
- support for the sustainability of fundamental initiatives, such as RedClara;
- help build political support and provide advice for these kinds of sustainable activities; and
- help build relations with the private sector.

Session 2. Environmental and Geospatial Data and Metadata

Antônio Miguel Monteiro from INPE, Brazil chaired this meeting that had Raed Sharif, from Syracuse University, USA as rapporteur. The following presentations were given prior to discussions:
- Silvana Amaral (INPE - GEOMA, Brazil): *Environmental & Geospatial Data for modeling processes in the Amazon Region. Challenges, barriers and goals over studies of Biodiversity and Land Use and Land Cover Change at GEOMA Project*,[27]

---

[27] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.2/samaral.ppt

- Puneet Kishor (University of Wisconsin, USA): *Open GeoSpatial Data: Difficult but Necessary*[28]*,*
- Holm Tiessen (Inter-American Institute for Global Change Research - IAI): *Data Needs for Adaptation to Global Change*[29]
- Luiz Augusto Machado (Division of Environmental Satellites-CPTEC/INPE, Brazil): *South America Remote Sensing Data Integration Based in Web GIS for Environmental Applications and Severe Storm Warning*[30]

Silvana Amaral presented the GEOMA project, a cooperative network of researchers whose objective is to develop environmental models to evaluate and forecast sustainability scenarios for the Amazon region. The network was created to help overcome environmental problems of the region and to disseminate scientific information and exchange knowledge among different research institutes. Challenges, barriers and goals faced by GEOMA's research team in modeling biodiversity and forest conversion associated to human processes of land occupation were presented. An initial database at a regional scale, including high-resolution remote sensing imagery, and some valuable models and prospective scenarios of deforestation are some of the results of the network. The presentation was concluded with a reminder that there are many important issues related to environmental and geospatial data that need to be seriously considered such as data quality, metadata, and scalability.

Puneet Kishor presented major legal, technical, and cultural challenges to sharing geospatial data. Legal challenges relate to contracts and data licenses necessary to attend the needs of both creators and data users. Technical challenges result from incompatible data formats and the complexity of managing both geometry and attributes that can make packaging and discovering geospatial data challenging. Cultural challenges stem from typically tradition-bound governmental agencies entrusted with the task of creating and managing geospatial data. In the final analysis, cultural factors can be substantially more important than legal or technological factors in determining spatial data sharing. The presentation was concluded with examples of projects and cases integrating biodiversity and environmental health information with geospatial information.

Holm Tiessen, talked about data needs for adaptation to global change. Global change occurs in time over space with associated risks, vulnerabilities, and opportunities. Global change science is driven by societal concerns and has an intrinsic requirement to communicate science output to society. The complexity of global change generates requirements for spatially addressing and integrating a multiplicity of data with various formats, granularities, scales and resolutions from many disciplines. This raises new issues of data management that need to be resolved in order to turn data into effective information for policy and decision-making. Tools that are needed in this context are databases that allow unique area-referencing of multiple data and information, and user interfaces that permit game-like exploration of decision options based on indicators derived from modeling interactions of different kinds of information. An effort is needed not only to manage data, but to seamlessly link data to knowledge and its applications. Within this context, the IAI addresses issues such as: climate change and variability; ecosystems, biodiversity, land use and land cover, and water resources; human dimensions and policy implications; risk, exposure, and vulnerability; regional analyses; and science, communication, and decision making.

The main goals of the Institute are to: (i) use conservation biology, ecology, biogeochemistry, remote sensing, sociology, anthropology, policy analysis, agriculture, and forestry to develop comprehensive understanding of environmental change, (ii) understand and predict the behavior of ecosystems and the displacement of their boundaries under environmental change, (iii) establish links between biodiversity, ecosystem functions, and ecosystem services, and (iv) characterize patterns and drivers of land use change.

Luiz Augusto Toledo Machado stressed the fact that South America has large volumes of environmental data that are not integrated in an open system. Brazil, for example, has around 23 weather radars but this information is not available to the large user community like Civil Defense or emergency management groups. Many other kinds of data like lightning, nowcasting models

---

[28] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.2/pkishor.ppt

[29] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.2/htiessen.ppt

[30] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.2/lmachado.ppt

and weather stations are recorded but the information is not available in real time and cannot be easily interpreted and understood. This presentation described the effort to integrate remote sensing information from different agencies and satellites through SIGMA (Sistema de Informação Geográfica Aplicado ao Meio Ambiente - Geographic Information System Applied to the Environment), a useful tool to interpret and integrate environmental information. The GEONETcast[31] system combined with SIGMA was also shown to demonstrate how this information could be available to a large community.

SIGMA is a web based Geographic Information System currently being developed and running at the Weather Forecast and Climate Studies Centre (CPTEC) of the Brazilian Institute for Space Research (INPE). SIGMA is totally dedicated to environmental applications and was designed to integrate different types of data related to the environment. SIGMA assimilates meteorological satellite data generated at CPTEC in real time, as well as weather radar, lightning and rain gauge data and convection nowcasting. Satellite data consist of GOES, MSG, NPOESS, MODIS and derived products, such as precipitation, surface temperature, lightning activity, fire occurrences, NDVI, ultra violet, ozone concentration maps, etc. SIGMA incorporates basic functions like zoom, pan, query builder, measuring tools and others to help users in visualizing and carrying out spatial analysis. Complementary layers such as road maps, cities, counties, rivers, location of airports and a Landsat mosaic are also available from the SIGMA. Digital Elevation Model derived from the Shuttle Radar Topography Mission (SRTM).

All this information is available on the Web in real time. Efforts should be undertaken to integrate more information such as 10 Amazon weather radars and other South American radars and weather stations. The main goal for the future is to prepare an open SIGMA version to work connected to the GEONET to improve the use of environmental data.

Discussion

After the presentations the group discussed a number of questions and topics that follow.

3. What are the current challenges and barriers to provide open and permanent access to environmental and geospatial data and metadata?

Different levels and categories of challenges and barriers to open and permanent access to Environmental and Geospatial Data and Metadata were identified. The first to be considered was *legal issues*. Participants indicated that there is lack of knowledge about intellectual property and copyright and instruments such as licenses and contracts. Initiatives such as Creative Commons seem to be adequate for articles but issues concerning data and databases must be further addressed. As to *policy*, there is a need for clear data policy at governmental, institutional and personal levels. Sponsorship restrictions on projects and scientists were also identified as barriers to data sharing. *Economic issues* such as "who will pay for accessing the data?" and "commercial versus non-commercial use" were also noted as a challenge when planning a program or project and an important barrier to open access. *Social and cultural barriers* such as technical language and communication problems between technical fields (i.e., information technology-related fields and other scientific fields) and barriers concerning natural language need to be overcome. Finally, legitimate *political issues*, such as military use of civil data, also need to be accommodated.

Concluding this discussion, a number of technical challenges and barriers were listed. These include: Interoperability; open standards and formats; insufficient infrastructure for real time data; metadata and scales; the need for more easily usable software and applications; storage and bandwidth; rights and privacy concerns; latency; and, quality, security and integrity of data.

4. What are some of the most promising existing models or mechanisms for providing such open access to health information, data, and metadata?

The group considered that it would be important to establish a forum to discuss issues related to open access and their importance to global change issues. There is a need for a public license for geospatial data and for educational toolkits about Creative Commons and other related issues for international funding agencies and global programs to be able to put these solutions into practice. Foundations should be convinced of the importance of open access to and archiving of project

---

[31] GEONETCast is a global network of satellite based data dissemination systems providing environmental data to a world-wide user community

data that they fund. The group considered it important to begin with the foundations as an example before reaching governments and private sectors. All agreed that sessions about Creative Commons and intellectual property rights should be promoted at academic, practitioner, government, and international conferences and meetings.

5. **How can such new or improved initiatives specifically be implemented and who should be involved?**

One recommendation was to disseminate the discussions held at this workshop and proposed recommendations to funding agencies at their next meetings and to try to involve young scientists, not only to help promote a cultural change but to promote more academic research in this field. The publication of best practices and case studies on data sharing related to these fields were also recommended. All actions suggested should be considered tasks for all to work on individually and as an institution.

The following parties where identified as potential stakeholders in the implementation of the above ideas: Global Programs; ICSU Regional Offices; CODATA; Science Academies; IAI; IDRC, UNDP, WB; IDB; CI; EU; GSDI; local country actors; and universities.

Session 3. Biodiversity Data and Metadata

Andrea Ferreira Portela Nunes from the Ministry of Science and Technology of Brazil chaired this meeting that had Bonnie Carroll, USA as rapporteur. The following presentations were given prior to discussions:

- Beatriz Torres (Outreach and Capacity Building, GBIF): *Promoting open access to biodiversity data and metadata: the GBIF approach*[32]
- Dora Canhos (CRIA, Brazil): *Open access to biodiversity data: the speciesLink experience*[33]
- Célio Magalhães (INPA, Brazil): *PPBio - Program for Planned Biodiversity Studies*[34]
- Victor Miyakawa, Siamazonia (IIAP, Peru): *SIAMAZONIA: A response to the challenge of managing and effectively using biodiversity information in the Peruvian Amazon*[35]
- Ximena Franco (Humboldt Institute, Colombia): *Biodiversity Information in Colombia: National and Regional Perspectives*[36]

Beatriz Torres indicated as main challenges and barriers to providing permanent open access to S&T data and information the amount and the complexity of data. There are over 300 years of work in identifying and naming species while less then 5% of data held in biological collections are digitized. Cultural barriers also represent a challenge to making data accessible. Convincing data providers to include metadata and to adopt standards are two examples. Other barriers worthy of mentioning are the lack of reward systems that recognize the effort to make data available; legal issues such as proprietary systems and intellectual property rights; restrictive policies or the lack of open access policies; and technological barriers such as interoperability.

Beatriz then presented GBIF as a promising model for providing access to biodiversity data, highlighting its capacity to establish partnerships. As GBIF is a gateway and does not own data, any kind of restriction must be determined by the data provider. Whatever is served to GBIF is made freely and openly available. An important characteristic of the GBIF approach is the establishment of clear rules and a working framework with partners. A number of tools such as monitoring species loss through time and ecological niche modeling were shown as examples of data use.

---

[32] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.3/btorres.ppt

[33] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.3/dcanhos.ppt

[34] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.3/cmagalhaes.ppt

[35] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.3/vmiyakawa.ppt

[36] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.3/xfranco.ppt

Dora Canhos presented CRIA's experience with the speciesLink network in Brazil[37]. In order to promote data sharing it is necessary to overcome cultural barriers, technical limitations, and legal impediments. But promoting data sharing doesn't only depend on the "will" to share data. It must be organized through the use of data models, standards, and protocols, it must be feasible or "doable" and it must be planned, with adequate resources, expertise, and infrastructure.

The challenge to implement the speciesLink network was to develop a distributed system without changing the management system in use and at the same time guarantee that biological collections have full control over their data, deciding what could be shared in an open network. Another challenge was how to deal with different levels of infrastructure and local capacity. Some collections have full Internet capabilities, use relational databases, and have a very high percentage of digitized data. Others have poor connectivity, use spreadsheets, and do not have any local computational support. The network to be developed had to attend all requirements and needs.

CRIA adopted DiGIR (Distributed Generic Information Retrieval) as protocol and DarwinCore as data model. A third component that was introduced in the architecture was the software "spLinker" that is responsible for mapping data fields, creating filters for sensitive data and controlling data flow (input, updating, and deleting) between collections and cache nodes. Besides building a distributed information network to make specimen data held at biological collections openly and freely accessible, a number of tools were developed for both data providers and users. Data providers can benefit from participating in the network by having their data automatically analyzed to find possible errors and to see the profile of their collections based on the data that is on-line. Data users have different options including output for the data retrieved (XML, Excel, HTML, visualization on a map) and a framework for ecological niche modeling[38]. CRIA set a number of indicators such as number of records per state and per family, proportion of digitized, georeferenced and non digitized records, etc. With the participation of more collections, these indicators may help identify information gaps (taxonomic and geographic) and plan further digitization programs.

Célio Magalhães presented the Program for Planned Biodiversity Studies (PPBio) established in 2004 by the Ministry of Science and Technology (MCT). MCT understands that biodiversity studies must be integrated and conducted over large areas, and that scientific information generated is important for education, environmental modeling and to support public policies. The program was conceived in accordance with the principles of the Convention of Biological Diversity and the directives of Brazil's National Biodiversity Policy. PPBio is part of the federal government's Multi-year Plan and aims at efficiently expanding and managing biodiversity knowledge. It has a national scope, but in the first two stages priority was given to the Amazon and Semi-Arid Regions. Specific objectives are to promote: (i) development and maintenance of inventory networks; (ii) maintenance, expansion and digitization of national biological collections; (iii) research and development on biodiversity thematic fields; and (iv) develop strategic actions for biodiversity research policies. The program's main activities include surveys, biological collections, commercial applications, strengthening regional centers, and capacity building.

The policy is to make metadata immediately available so that other researchers and managers know what is being collected, when, and where. As to data access, the principle is that data collected in public lands, or by government employees, or using infrastructure provided by government funding are public property and must be made available in a timely fashion. In accordance to PPBio's data policy, researchers have a maximum period of 2 years for the data to become public and be made freely available on the web. The Biological Collection component, besides improving the collections' holdings and training human resources, aims at improving data management, data quality, and in integrating biodiversity information systems. The next biome to be integrated to the program is the Atlantic Rain Forest.

Victor E. Miyakawa presented SIAMAZONIA, a response to the challenge of managing and effectively using biodiversity information in the Peruvian Amazon. Presently, there are several decentralized systems that are directly or indirectly handling information about biodiversity. Most of these initiatives are global or regional (GBIF, Species 2000, Manis, IABIN, etc.), with national nodes or hubs which assemble and share information locally. However, national providers (nodes)

---

[37] see splink.cria.org.br/

[38] see openModeller available at openmodeller.sourceforge.net/

not only share information to fulfill the common objectives outlined by the global and regional initiatives, but also have the need to fulfill national demands related to biodiversity information. In order to meet these challenging demands, local nodes need to understand and engage in processes that involve the implementation of strategies for using data for decision support systems and policy making processes.

Global strategies like GBIF have significantly helped not only to increase awareness about the importance of biodiversity information, but also with practical methodologies and tools necessary to handle, provide access and disseminate biological diversity information. However, due to the complexity and range of action of initiatives such as GBIF, among others, full use of these data in local environments is still a challenge. In Peru, more specifically in the Peruvian Amazon, (which comprises 69% of the country) managing and using this information is a must.

SIAMAZONIA (www.siamazonia.org.pe), the Peruvian Amazonian Biodiversity Information System, is a mechanism oriented to make full use of this type of information. Its impacts in Peru range from developing capacities in the biodiversity informatics fields to generating thematic systems that make use of the information managed by SIAMAZONIA. The implementation of this information system has increased confidence in sharing information through a series of practical strategies and mechanisms. These have resulted in an increasing number of generators/custodians of biodiversity information which have fully accepted the challenge of managing, sharing and disseminate biodiversity information related to the Peruvian Amazon.

Ximena Franco presented national and regional perspectives concerning biodiversity information in Colombia. The context of information sharing in Colombia is very complex. Data providers include governmental research institutes, universities, academic research centers, NGOs, biological collections, scientific societies and networks. Data users include the national system of protected areas, regional and local environmental authorities, national authorities, and the general public. The Alexander von Humboldt Biological Resources Research Institute was created in 1993 to promote, coordinate and undertake research that contributes to the knowledge, conservation and sustainable use of biodiversity in Colombia. One of its functions is to implement an information system as a supporting tool for the National Biodiversity Inventory. Its strategy is to encourage alliances with agencies, researchers, and authorities to build a National Information Network: called the Biodiversity Information System of Colombia (SiB)[39]. The Humboldt Institute through SiB is providing tools to facilitate access, sharing and use of biodiversity data. They are also promoting a cultural change towards open sharing of data. Important issues include: confidence and transparency; intellectual property rights; clear rules and policies for data sharing; and, distributed ownership (equal opportunities for contribution and participation). The Biodiversity Information System (SiB) of Colombia is a national alliance developed to facilitate data and information management. It seeks to efficiently support research, education and decision making processes related to knowledge, conservation and sustainable use of the biodiversity of Colombia.

More than an information facility, SiB is a knowledge community, supporting not only research and decision making processes, but also providing tools for the scientific community and decision makers. The aim is to strengthen decision makers' capacities in handling and using information and tools.

Discussion

The presentations were followed by discussions that focused on *current challenges and barriers to providing open and permanent access to biodiversity data and metadata.*

Different levels and categories of challenges and barriers to open and permanent access to environmental and geospatial data and metadata were identified. There was consensus as to the complexity of collecting and managing biodiversity data and the huge task of digitizing biodiversity data for species and specimens of which less than 5% are currently digital. It was also recognized that there is a diversity of stakeholders (e.g., the indigenous populations' cultures and languages) and a need to integrate diverse data over large areas. Information gaps were recognized as many areas and taxa remain unexplored. Therefore, in addition to digitizing legacy data it is important to carry out new surveys.

---

[39] see www.siac.net.co/

Lack of strong political, technical and commercial uses of information was identified. There is a need to effectively use biodiversity information and to link scientific and technical systems to the environmental system for land use planning, monitoring impacts, for managers/decision making.

Data providers' resistance to data sharing and to the use of standards was acknowledged, both in collecting techniques and in adopting biological protocols and metadata standards. It is necessary to connect needs with research processes through information capacity. Data providers must have full control over their data, to assure necessary filtering of sensitive data and respect for traditional knowledge. But there is a need for cultural change towards the open and free sharing of data in a timely manner.

Sustainability was also stressed as a major challenge. There is a need for long-term stable financial support to collections and information systems (not just project based). Besides financial constraints, there is a need for capacity building for human resources. Long term data preservation policies should also be addressed. Clear rules for data sharing with consistency and understanding of intellectual property rights are fundamental, as are clear policies concerning biodiversity research and the generated data.

There is also a need for a reward system or incentives for publishing and sharing databases or information just as is done with publications. Another barrier mentioned was lack of universal Internet access as only a small percentage of the population in most Latin American countries has access.

The group identified a number of positive attributes of the activities that were presented. GBIF and CRIA have a good set of data policies with their providers. Data belongs to the providers with full attribution, metadata are in place and sensitive data are held back by providers. GBIF has data sharing agreements and a formal open access policy. Both systems aim at developing local capacity and are inclusive. It was the opinion of all that decentralization with data providers having full control over their data builds confidence.

PPBio presented an interesting program for integrated biodiversity surveys and a policy for data sharing established upfront. This should be the rule for all projects that generate biodiversity data that must be shared.

Both SIAMAZONIA in Peru and Colombia's Biodiversity Information System (SiB) are national efforts and are also inclusive and decentralized. SIAMAZONIA includes traditional means of communication not only the Internet, a feature that will be interesting to follow. SIAMAZONIA is also developing practical uses of information aiming at involving the private sector to promote bio-commerce.

The following potential follow-up projects were discussed:

6. Cooperate on the development of the Amazon Basin Biodiversity Information Facility

GBIF nodes of Peru and Colombia together with information centers and biological collections from Brazil, Cayenne, Ecuador, Bolivia, and Venezuela submitted a proposal for a GBIF campaign to create a network of cooperating institutions to share and disseminate species and specimen data and information about the Amazon Region. This initiative will be country driven and aims at providing a free and open data and information infra-structure to improve research, policies, and decision-making. Products will also include a checklist of Amazonian Flora and Fish

7. Share the lessons learned from successful biodiversity informatics centers/projects

Ways to share these lessons learned on a continuous basis should be developed.

8. Prove the value (economic, cultural, social, decision making) of biodiversity information

The idea is to work with the scientific community to publicize the value of access to scientific data and connect it to key issues (e.g., the 2010 challenge of species decline; or the consequences of climate change). Another suggestion was to demonstrate the return on investment through case studies and to work to better understand the impact of access to information by determining valid metrics and comparing the Internet with different modes of communication.

9. Develop and promote a reward system approach for database management and publication

CODATA might consider a Working Group or Task Group on this topic. ABBIF, GBIF, and Brazil (CGEE) could participate. The activity could also examine the role of intellectual property rights in the reward system.

Session 4. Scientific, Technical and Medical Journals

Rogério Meneghini from BIREME, Brazil chaired this meeting and Paul Uhlir, NAS, USA, was rapporteur. The following presentations were given prior to discussions:

- Abel Packer (SCIELO, Brazil): *SciELO as a model for scientific communication in developing countries*[40]
- Leslie Chan (University of Toronto, Bioline International, Canada): *Collaborative approach to open access: Experience from Bioline International*[41]
- Tom Goehl (publishing consultant, USA): *Adapting the African Journal Partnership Concept to Latin American Countries*[42]
- Clemente Forero-Pineda (Universidad de los Andes, Colombia): *Citation indexing systems and the future of scientific publications in Latin America*[43]

Abel Packer presented SciELO, which was created 10 years ago with the objective to help researchers in Spanish- and Portuguese-speaking developing countries in becoming active participants in the international system of scientific publication and communication. The international journal system is dominated by ISI evaluation and impact factors, and historically, the whole system has been skewed to favor established OECD countries' journals, most of which restrict access primarily to those able to pay high subscription fees. SciELO has resulted in vastly increased visibility and breadth of accessibility of Latin American publications through online open access, and in a significantly higher impact factor for many journals and individual articles. SciELO also maximizes links to other open scientific communication outlets such as Medline and increases access speed. The SciELO mechanism is both an online network and a repository. Economic sustainability is addressed primarily as a public good activity by governmental S&T agencies, as well as participating academic and scientific societies, sponsors, authors, private publishers and brokers.

Leslie Chan indicated that Bioline International has goals and effects similar to SciELO, but unlike SciELO, which is organized at the national and regional levels, Bioline is more of a grassroots project for individual journal publishers. Bioline provides a platform for open, low-cost, volunteer and, collaborative approaches to the formation and implementation of peer-reviewed journals in developing countries. The project supports the development of tools for OA publishing, advocacy, and research on the benefits of OA publishing. Bioline also serves as an open repository, which is hosted by CRIA. Bioline infrastructure funding is sustained though a mix of core institutional support, grants, in-kind support, and many volunteers. An exclusive focus on impact factors by authors at the expense of accessibility may be seen as an ethical issue. The funding of participating journals is based on a mix of sources at the more local level.

Dr. Chan suggested further actions that might be considered by OA publishing:

- new value-added services with OA databases and open standards could be developed;
- open linking to open databases and other literature should be promoted;
- more comprehensive measures of scientific, economic, and social impacts of the literature, in addition to the ISI citation index should be developed; and
- funding agencies should be actively informed about the benefits of OA publishing.

---

[40] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.4/apacker.ppt

[41] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.4/lchan.ppt

[42] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.4/tgoehl.ppt

[43] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_09-05/Session7/panel7.4/cpineda.ppt

Tom Goehl informed the session participants that the Fogarty International Center and the National Library of Medicine, both of the U.S. National Institutes of Health (NIH), developed a program in 2003-2004 to implement editorial partnerships in Africa. The initial focus was on four African journals in health research that also had access to sufficient bandwidth. The mission was to promote and improve the publication and dissemination of health research information in sub-Saharan Africa. The Council of Science Editors provided links to editors at top health journals in the United States and the United Kingdom, including African Health Sciences with the British Medical Journal, Ghana Medical Journal with The Lancet, Malawi Medical Journal with the Journal of the American Medical Association, and Mali Medical Journal with Environmental Health Perspectives and the American Journal of Public Health. The tasks performed under this project include identifying equipment and facility requirements and editorial needs, providing hardware, software, author and reviewer training and business plan support, managing editor and business manager support, website support for OA, internships, methods of writing review articles, evaluation metrics, and administrative support. The total cost for all four journal projects has been about 75 thousand USD per year.

Clemente Forero-Pineda spoke about citation indexing systems and the future of scientific publications in Latin America, indicating that it is important to recognize that access to scientific information is not sufficient. Users require a measure of quality. One such measure has been citation indexing, although other measures are also needed. When Eugene Garfield conceived citation indexing in the 1950s, he proposed a set of principles and a utopian agenda. The basic principles have been implemented, but not the utopian agenda: polyglot indexing, regional citation systems, and Robert K. Merton's additional arguments for open access to scientific information. Fifty years of citation indexing has induced profound transformations of science in Latin America, which has been documented by Prof. Forero in a survey of 92 journals and 61 leading scientists. A measurable divide in local scientific communities has developed and Spanish and Portuguese journals are under-represented in ISI and Scopus. SciELO has helped increase visibility of Latin American journals, but a longer term goal would be the federated networking of world citation indexing systems. A logical first step would be decentralized networking of open access indexing systems.

Current challenges and barriers to providing open and permanent access to journals

There is confusion about ownership and assignment of copyrights by authors and research institutions. Some publishers use copyright as a tool against the broader interests of the publicly funded research community. The exclusive assignment of authors' copyrights to publishers blocks access and reuse of journal articles, especially through text mining and other automated knowledge extraction applications. These technological benefits from open access online need to be better exploited, while the technological barriers to such automated applications from digital rights management software need to be mitigated.

The potential economic benefits and network effects of online scientific data and information also are poorly understood and this lack of understanding undermines support for open access publishing and deposits, which would enhance the value of this material through much greater availability and use. The potential social benefits of open access are known and appreciated even less than the economic effects. Additionally, the financial sustainability of open access models needs to be better understood and addressed for the long term.

Discussion

What are some potential realistic cooperative activities in Latin America to help address the challenges and barriers that have been identified? The following specific activities were developed during this session.

Enhance open access publication services. Various methods for promoting open access should be supported within the research community. For example, linking to open databases and other literature with enhanced functionalities is crucial (e.g., species links, adding geospatial dimensions, automated translation). The broader distribution of open-source value-added tools would also be useful.

This could be accomplished by inventorying existing open-source tools and publicizing and disseminating existing software programs for enhancing the value and functionality of links between the literature and the underlying data. A pilot project could be initiated in the area of public health publications in Latin America with potential support by UN GAID and PAHO.

Develop more comprehensive measures of impact and value of open access. A typology of impacts that moves beyond mere citation indices needs to be developed and then utilized. Research impacts are more than ISI (e.g., Google PageRank). Social impacts would focus on uses by non-researchers, social benefits (e.g., health benefits from access to biomedical and health journals; use of open access publications in education; use in the formation of public policy, and so on). Economic impacts could examine R&D development and economic returns. A comprehensive methodology would include both empirical statistics as well as more subjective approaches.

For example, the U.S. CODATA is planning such a methodological survey in a workshop with the OECD in 2008 regarding public sector information (PSI), a related information area. This could be followed up with assessing the relative benefits of open access and restricted publishing of scientific data and literature would have similar issues and applications and this could be followed up. The Brazilian CODATA Committee and the Brazilian Academy of Sciences could follow up on this issue as well. Another potential project is to start implementing indexing methods with open access literature published under Creative Commons licenses, for example in tropical diseases. The Science Commons is interested in doing this using SciELO and Bioline literature. Funding for such activities might be solicited from the IDRC, the Open Society Institute, and some U.S. science agencies.

Establish an international, polyglot indexing system through a federated network of open access indexing systems. Can the existing open access and subscription publishers and the stakeholders in the system move toward this vision? A meeting, or series of meetings, could be convened with all the indexing services, such as ISI, Scopus, Medline, and regional ones, such as Pascal, LatIndex, SciELO, and the other stakeholders in the research system (e.g., open access journal editors, Academies, societies, and government research organizations). The project could begin small, with the involvement of only the indexing services, and then expand to the other stakeholders. The initial objective would be to identify the key issues, opportunities, and challenges. Clemente Forero-Pineda agreed to develop a proposal for such an activity.

Establish a Latin American partnership for [public] health journals. The same or similar methodology as for the African Journal Partnership could be adopted. However, a clear purpose needs to be established and criteria for selecting journals in less developed countries in Latin America need to be developed (e.g., some Andean or Central American countries). The situation in Latin America is different from Africa so this needs to be expressly considered. Tom Goehl is willing to draft a proposal. PAHO, BIREME, and SciELO could be partners with the NIH in the implementation of such a project.

PART III. CROSSCUTTING BREAKOUT SESSIONS

In addition to the thematic breakout sessions, the workshop included sessions to facilitate cross-disciplinary discussion and proposals. Participants were requested to respond to the following questions, based in the results and presentations from the previous sessions:

> What are some potential realistic projects or collaborations to help address the challenges and barriers that have been identified?
>
> How can the new or improved initiatives specifically be implemented?

Session 1. Integration of Geospatial, Socio-Economic and Biomedical Data for Poverty Reduction and Sustainable Development

Participants were asked to focus on initiatives, models, methods, challenges, best practices and lessons learned for integrating geospatial, socio-economic and biomedical data for poverty reduction.

This session was chaired by Luiz Henrique Proença (IPEA, Brazil) and Raed Sharif (Syracuse University, USA) was the rapporteur. Presentations included:

- Christovam Barcellos (Fiocruz, Brazil*): Integration of environmental, social and health data using GIS: Lessons learned from three disease outbreaks investigations in rural areas*
- Carlos Castillo-Salgado (PAHO): *Uses of Geographic Information Systems for Global Public Health. The GIS-EPI Project*
- Marjorie McGuirk (NOAA, USA): *Integrating Environmental and Social Data*
- Paul Uhlir (The U.S. National Academies) and Raed M. Sharif (The U.S. National Academies and Syracuse University): *Geospatial Data for Health and Environmental Applications in Latin America: A Focus on Poverty Reduction and Sustainable Development*

Integration of environmental, social and health data using GIS: Lessons learned from three disease outbreaks investigations in rural areas.[44] By Christovam Barcellos, Fiocruz, Brazil

Three recent experiences of zoonotic diseases epidemiological investigations in Brazil were described in this presentation: Hantaviruses in the southern Brazilian region, cutaneous leishmaniasis in a rural settlement, and toxoplasmosis in an urban area. The role of spatial data acquisition and analysis is reinforced, considering the intrinsic complexity of environmental and social factors affecting disease transmission. Disease transmission, particularly during outbreak situations, is a consequence of close relations between human and animal populations with the environment, i.e., the ecology of the animals involved in the transmission cycle and the human work in those areas. Spatial analysis allowed in all cases to characterize the ecology landscape in which outbreaks occur. Simple and available spatial analysis tools were used to complement epidemiological field investigations. Landsat satellite image (Embrapa, www.cdbrasil.cnpm.embrapa.br) was used to characterize land use and suspected risk sources. Spatial data were analyzed using Terra software (www.dpi.inpe.br).

Health surveillance in Brazil is undergoing a decentralization process, according to which, different responsibilities are required from each health institution, from federal level to county health secretaries, including local Non-Governmental Organizations. Investigations involved the multiple levels of health agencies. This practice assures the continuity and transparency of health prevention and promotion actions. These institutions must be integrated to assure rapid and adequate information flux.

The availability of data, as well as skilled use of spatial analysis tools, contributed in this process, allowing rapid and low cost assessment of environmental risks. Lessons learned from these experiences involve the development of spatial analysis tools, low cost technology appropriation by health professionals and the structure of health surveillance system in large and unequal countries such as Brazil. This system is based on local health surveillance and on people who must be able to acquire and use the available spatial data and analytical tools to detect and notify the public about disease clusters. Considering information availability in Brazil, these practices could be incorporated to routine of health services using low cost digital mapping in localities with modest infrastructure, together with free information and software.

Uses of Geographic Information Systems for Global Public Health. The GIS-EPI Project.[45]
By: Carlos Castillo-Salgado, PAHO

PAHO is the oldest international public health agency in the world agency with 100 years of experience working to improve health and living standards of the people of the Americas. It is part of the United Nations system, serving as the Regional Office for the Americas of the World Health Organization, and as the health organization of the Inter-American System. The Pan American Sanitary Bureau (PASB) is the Secretariat of the PAHO. The Bureau is committed to providing technical support and leadership to PAHO Member States as they pursue their goal of Health for All and the values therein. One of the important projects at PAHO is the GIS-EPI project.

The functionalities, services and millstones of GIS-EPI, which is a comprehensive platform for developing country and institutional capacity in the Americas for the use and application of GIS in Public Health, were presented and briefly discussed. The project includes training in GIS/Public Health, joint development of applications for global health mapping, epidemiological surveillance

---

[44] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.1/cbarcellos.ppt

[45] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.1/ccastillo.ppt

and health situation analysis. It also facilitates the dissemination of best practices, experiences, scenarios and problem-solving applications in Public Health.

Integrating Environmental and Social Data.[46] By Marjorie McGuirk, NOAA, USA

Government agencies and researchers that collect and use weather data for social purposes have long experience with both the problems of integrating large, diverse, and distributed data and with the potential solutions to these problems. Other multi-disciplinary research fields share the problems. Furthermore, solutions in one such area can be helpful in others. This presentation concisely summarized many key problems faced in integrating environmental and social data for a variety of purposes, including open and permanent access to the data. The problems include physical challenges of large data volumes, interdisciplinary language differences, data format differences among different scientific fields, to name a few.

These same agencies are also engaged in developing the infrastructures and work practices required to provide the access needed. The Group on Earth Observing Systems of Systems (GEOSS) has a series of cooperative programs that include Latin American countries. The World Meteorological Organization is promoting open access to data for global exchange. Geospatial information research is another international field that is developing working data sharing mechanisms. NOAA is working with Uruguayan Meteorological Service to share current and archived observations. And in April 2007 NOAA's GOES-10 satellite was repositioned over South America.

Geospatial Data for Health and Environmental Applications in Latin America: A Focus on Poverty Reduction and Sustainable Development. By: Paul F. Uhlir, The U.S. National Academies and Raed M. Sharif, The U.S. National Academies and Syracuse University

A proposed collaboration between the US National academies, the Inter-American Network of Academies of Sciences and the Pan-American Institute for Geography and History in the area of the use of GIS for health and environmental application for poverty reduction and sustainable development was presented and briefly discussed. This proposed collaboration is to: (i) identify the types and volumes of publicly funded geospatial data pertinent to the coverage of the region that are produced in Latin America (LA) or by sources outside LA, and are maintained and used in LA; (ii) describe the laws, policies, and other key factors that affect access to and use of those data; (iii) describe and evaluate the current status of access to and use of those data, including the application of digital networks for such purposes, and characterize the uses and the users of such data in the health and environment fields, and (iv) organize a major workshop in LA to identify and discuss the impact of access to and use of publicly funded geospatial data on the development, dissemination, and use of health and environment-related GIS tools and applications, and the potential effects on poverty reduction and the broader sustainable development efforts in the region.

Discussion

Legal, political, economic, social and cultural challenges and barriers in integrating geospatial, socio-economic and biomedical data for poverty reduction and sustainable development were identified.

Legal. Geographic information systems (GIS) have very few regulations concerning privacy and security. Therefore, ownership protection becomes more problematic for both data providers and users as licenses, contracts and copyrights issues are still not well understood and have not been widely implemented.

Political. In most countries, GIS data are purposely not consistent (e.g., geographic boundaries). Primary producers of geospatial data are governments, and getting accurate GIS data may be challenging at times. There is a need for policies on open access and archiving at universities and research centers. There is also a need for training and coordination across political boundaries. In some countries there is also the problem of "politicization" of health data.

Economic. Obtaining GIS-related software and hardware and training on these tools at a local level can be challenging and normally is very expensive. Data exchange is very power intensive, which also makes it very expensive.

---

[46] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.1/mmcguirk.ppt

Social and cultural. Important societal issues often require data/observation from a variety of observation systems. Most of the time, collecting, obtaining, and analyzing these data is very challenging. Problems such as incompatible format between science disciplines (e.g., naming standards, location standards, and format) are quite common. There are also natural language challenges and barriers. Health also involves issues related to privacy of individual patients' data.

Technical. The use of second (and third) level boundaries (state and municipality) is very important for health and GIS and here again there may be problems with metadata and coding. Challenges include linking health data with other fields' data such as environmental and social information systems.

The following projects and ideas, together with potential partners and stakeholders were identified and briefly discussed:

10. Geospatial Data for Health and Environmental Applications in Latin America: A Focus on Poverty Reduction and Sustainable Development.

Potential partners and stakeholders include the U.S. National Academies to take the lead, the Inter-American Network of Academies of Sciences, and the Pan-American Institute for Geography and History (PAIGH).

11. Implement the 2nd level boundaries (PAHO will be distributing all 2nd level boundaries to all Latin America countries)

Potential partners and stakeholders include PAHO, to take the lead in drafting a proposal, WHO, UN, and national statistical institutes.

12. Develop methods and criteria to validate health data.

Potential partners and stakeholders include RIPSA (to take the lead), PAHO, and National Health Information Systems.

13. Identify and leverage existing ongoing initiatives at regional and national levels towards linking health information systems with other social and environmental data systems.

Potential partners and stakeholders include Brazilian and US Science Academies, CODATA, and GEOSS.

Session 2. Integration of Environmental, Geospatial and Biodiversity Data for Improved Research and

The panel focused on initiatives, models, methods, challenges, best practices, and lessons learned from integrating geospatial & biodiversity data for improved research and applications.

Antônio Mauro Saraiva, (Escola Politécnica of the State University of São Paulo, Brazil) chaired the session and Puneet Kishor (University of Wisconsin, USA) was rapporteur. Presentations included:

- Guido Gelli (Geo Sciences Director, IBGE): *Brazil Georeferenced Information System on the Brazilian Amazon.*
- Tim Sutton and Renato De Giovanni (CRIA, Brazil): *openModeller Desktop - A new approach to niche modeling*
- Marinez F. de Siqueira (CRIA), Angélica Giarolla (CPTEC/INPE), Lúcia G. Lohmann, (Biology Institute, State University of São Paulo, Brazil): *Environmental satellite data: Applications for the study of the physical environment and biodiversity*
- Oliver Phillips (RAINFOR/ATDN, UK) and Vanderlei Canhos (CRIA, Brazil): *Toward integrating three large and disparate networks and databases of Amazon tree biodiversity*

Georeferenced Information System on the Brazilian Amazon[47], *Presented by Guido Gelli, Geo Sciences Director, IBGE, Brazil*

---

[47] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.2/ggelli.ppt

The main objective of this project is the production of basic physical (geology, geomorphology and soils) and biological (vegetation) information of all of Brazil, in order to assist thematic studies, environmental diagnostic projects, production of indicators of sustainable development for environmental statistics and public policy. The project specifically aims to establish a national GIS database, containing information on four of its main ecological variables: subsoil (Geology), soil, relief, and vegetation, with homogeneous and regularly updated coverage at 1:250.000 scale. Information on biodiversity (fauna and flora), hydrology, and land use are also added this database. This is an ongoing activity, part of the basic objective of IBGE to ensure information and study of natural statistics, geography, cartography and demography necessary to understand the physical, economic, and social reality of the country, especially for planning and national security.

openModeller Desktop - A new approach to niche modeling[48], Tim Sutton, Renato De Giovanni, CRIA, Brazil (presented by Tim Sutton*)*

The openModeller project (see http://openmodeller.sf.net) provides a set of software libraries and applications for carrying out ecological niche modeling. Ecological niche modeling is based on the premise that one can compute the range of environmental preferences for an organism by extracting environmental values (temperature, precipitation etc) at points where the organism has been recorded (e.g. herbarium collections), and then extrapolate these preferences into a generic model that describes the areas where the organism is probably able to survive. Various software packages already exist for this purpose - each one presenting a new learning curve for the user and usually targeted to a single algorithm. Comparison between algorithms becomes difficult when models are produced by different software applications.

openModeller Desktop enables researchers to create, compare, and analyze fundamental niche models produced with different algorithms using a single application. openModeller Desktop is one of the available interfaces on top of the openModeller library. Features include:

- multi-platform (the same software runs on Windows, Mac OS X and GNU Linux).
- multiple algorithms supported (GARP, Climate Space Model, Bioclim, Environmental Distance, etc.).
- allows for local or remote (using Web Services) model execution.
- visualization and reporting for completed models.
- automated data retrieval from online databases.
- designed for conducting large experiments.

This presentation provided a visual tour of the openModeller Desktop application.

Environmental satellite data: Applications for the study of the physical environment and biodiversity[49] *Presented by Marinez F. de Siqueira, CRIA*

Scientific research is becoming more and more dependent on environmental satellite data of geostationary orbit (GOES and METEOSAT) and polar orbit (NOAA, TERRA and AQUA). The information received through these stations is pre-processed and stored in databases and the data are then used to generate information to monitor climate and development of new products. Specifically in the case of surface research, data on soil humidity, vegetation indexes and evapo-perspiration are of major importance for the evaluation of the energy exchanged between the surface and the atmosphere. Furthermore, this data provides critical information for studies on the physical environment such as biological species niche modeling, forecast of plants harvest and agricultural zoning, among others. In the specific case of biodiversity niche modeling, surface data has greatly improved the quality and applicability of the models generated. Another major contribution of remote sensing data relates to the fact that it can contribute data for areas that no longer have surface data. The presentation addressed the use of vegetation indices and soil humidity data to model the potential distribution of selection species in poorly collected areas of South America.

---

[48] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.2/tsutton.pdf

[49] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.2/msiqueira.ppt

Toward integrating three large and disparate networks and databases of Amazon tree biodiversity[50], Oliver Phillips, RAINFOR/ATDN, UK, Vanderlei Perez Canhos, CRIA, Brazil (*presented by Vanderlei Canhos*)

Increasingly ecologists are using large, international research networks to standardize field data protocols, and collate vegetation plot information to help address large-scale questions concerning the distribution, function, and changes in tropical biodiversity.

Within the Amazon region we propose to integrate three disparate but powerful sources of information:
- long-term monitoring plot records on growth, dynamics, biomass changes, and species changes, from >100 plots across Amazonia (RAINFOR - Red Amazonica de Inventarios Forestales)
- floristic data on diversity, composition, and relative dominance, from 700 plots across Amazonia (ATDN = Amazon Tropical Diversity Network)
- geo-referenced specimen data from localities across Amazonia (ABBIF = Amazon Basin Biodiversity Information Facility, networked links to regional and international herbaria)

The talk introduced the three networks, discussed their current strengths and weaknesses, and argued that their integration would enable development of powerful new capabilities. In particular, the integrated resources would allow us to (i) better model the current relationships of Amazon tree taxa to climate, and thus (ii) provide the means to identify which taxa are at risk from projected climate change scenarios, and (iii) help to develop a distributed biodiversity observation and monitoring system for the world's largest tropical forest region.

Session 3. Creating Permanent Information Commons for Science in Latin America

This session was chaired by Barbara Kirsop (Electronic Publishing Trust, UK). Paul Uhlir (National Academy of Sciences, USA) was rapporteur. Presentations included:
- Tom Moritz (Conservation Commons, USA): *Open Access and the Conservation Commons: putting biodiversity data and information to work*
- Hélio Kuramoto (IBICT, Brazil): *Scholarly Information: a new open access model for Brazil*
- John Wilbanks (Science Commons, USA): *Accelerating Scientific Research Cycles through an Information Commons*
- Luis Villarroel (Ministry of Education, Chile): *Building capacity in the Latin American region in intellectual property policy and open access mechanisms for scientific information.*
- Jorge Machado (USP, Brazil): *Access to Scientific Information in Brazil*

Open Access and the Conservation Commons: putting biodiversity data and information to work, Presented by Tom Moritz, Conservation Commons, USA[51]

Environmental degradation and species loss continue to accelerate. Solutions may be found to reverse these trends, but these will only be possible with comprehensive data, information and knowledge on the conservation and sustainable use of biodiversity. Accessing biodiversity data and sharing conservation knowledge are not simple tasks, however. Difficulties abound. Much of the data, information and knowledge require by conservationists are fragmented, difficult to find, or simply not accessible. This is considerably magnified in many developing countries where the consequences of under development and the "digital divide" present enormous challenges to the realization of successful conservation efforts on the ground.

Managing the vast amounts of data and information we generate daily related to biodiversity represents a huge challenge to conservation organizations. Equally important is the challenge of ensuring the availability of these assets on demand to decision makers, policy makers, and others in ways that are easily "fusible" into a variety of applications, visualization tools, and decision support systems. These global biodiversity data sets represent the foundations of our knowledge

---

[50] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.2/vcanhos.ppt

[51] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.3/tmoritz.ppt

related to life on this planet, and a significant contribution to ensuring long term sustainable development and human well being.

The Conservation Commons is a growing, global partnership addressing the challenges of fragmentation, access, and integration we face in the realm of biodiversity data and information. These impediments not only affect the success of conservation efforts, but also impinge on government and private sector decision making which directly (or indirectly) impacts biodiversity. The conservation community is recognizing the need to unlock biodiversity data and information, and put these assets to work for conservation. A growing international cooperative effort of conservation organizations, research agencies, scientific institutions, and corporations - the Conservation Commons - is thus working to remove barriers to access, and more effectively connect practitioners to data and information assets in four key areas:

- removing legal and institutional barriers to open access;
- improving data compatibility/integration standards;
- building on recent innovations in GIS for biodiversity data "discoverability" and "fusion";
- building the architecture for open access publishing, and electronic archiving.

The Conservation Commons has coalesced around an agreed set of principles, defining a commonly held vision for open access, shared responsibilities, and ensuring the most effective use of biodiversity data and information. Over 70 institutions have formally endorsed these principles since the Conservation Commons was launched at the IUCN World Congress in November, 2004 The purpose of the Conservation Commons is to ensure open access and fair use of data, information, and knowledge on the conservation of biodiversity.

Scholarly Information: a new open access model for Brazil, Presented by Hélio Kuramoto from IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia[52]

Access to scholarly information has become a challenge to developing countries like Brazil. The crisis affecting scientific journals has been discussed at length but, many times, its real dimensions are unknown. Besides the rapidly rising journal's subscription prices, there are other factors like those concerning copyright, latency between results and their actual publication, and restriction on what can be published and how it can be disseminated. IBICT has joined the movement of Open Access to Knowledge and has used OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) compliant software packages to store and disseminate scholarly information. The idea is to design and implement a national policy of open access to scholarly information.

Accelerating Scientific Research Cycles through an Information Commons, Presentation by John Wilbanks

One promise of the commons approach to scientific information is the acceleration of research through faster, more efficient knowledge distribution and re-use. The Science Commons uses a holistic approach to moving knowledge through the classic scientific research cycle using commons-based methodologies, with a particular focus on implementations in neuroscience. This cooperative activity is based on open-source knowledge management software and content platforms for life sciences research.

Access to Scientific Information in Brazil, Pablo Ortellado, Jorge Machado, Gisele Craveiro (*presentation by Jorge Machado*)

This research project seeks to assess the role of copyrights in restricting access to publicly funded scientific knowledge in Brazil. The research will analyze the production of scientific journals, scientific books, text books, and software in an attempt to map copyright policies, funding sources and the positions taken by the various actors in the journal, book, and software production chain.

The hypothesis to be tested is that most of the scientific production protected by copyrights is financed with public funding and does not adopt a policy of open access. Thus, the public pays twice for scientific knowledge: it pays to produce it in universities, research institutions and public funding agencies; and it pays to have access to the resulting journals, books, and software.

This study will be executed by the Research Group on Public Policies of Access to Information at the Faculty of Arts, Sciences and Humanities of the University of São Paulo (USP). This interdisciplinary group, comprised of young professors at one of the most important Brazilian

---

[52] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session10/Panel10.3/hkuramoto.ppt

universities, seeks to produce works that encompasses theoretical and empirical research, which contributes to public access to information and knowledge. A better understanding of the prevailing situation can help build the case for open access to policy makers and research managers. The USP will share the results and similar research could be undertaken in other countries.

Discussion

Current challenges and barriers to providing open and permanent access to data and information together with the legal, political, economic, and social aspects were discussed

Legal barriers include restrictive intellectual property (IP) legislation and regulations. Another problem that was identified is the lack of knowledge about common-use licensing, such as the licenses developed by the Creative Commons and the Science Commons. There is also a lack of understanding of open access processes and its value by the research community stakeholders, which was classified as a social barrier. There are many endemic misconceptions and fallacies about open archiving that need to be overcome.

Active lobbying by big publishers against open access was indicated as a political barrier, together with a lack of lobbying and organized promotion of open access by the research community.

Sustainability of permanent information commons is a legitimate question, especially in developing countries, and presents an important economic barrier and a challenge to overcome.

Major technological barriers such as commercial standards, proprietary formats, digital rights management technologies, and technological protection measures that obstruct open access were also mentioned.

As to potential realistic cooperative activities in Latin America to help address the challenges and barriers that have been identified, the following projects, ideas, and some potential partners and stakeholders were identified and discussed:

14. Promote Latin American Participation in the Conservation Commons

It is important to put the conservation community's knowledge resources into a commons environment, based on 3 principles: open access, mutual benefit, and responsible use. The Conservation Commons is being organized under the IUCN and most participation is at the international level. There are many participants already, but it would be useful to add national and regional organizations from Latin America. The proposed action is to identify the organizations that would have a likely interest in this and then to contact them.

15. Promote the Ibero-Latin American Open Access Repositories (ILOAR).

ILOAR is establishing open archives for S&T literature and data in Latin America, building on successful models and open-source tools. There is a need to strengthen "human interoperability" and convince all segments of the research community to comply with open access to information.

The proposed open archives model is based on the Open Archives Initiative. Many open-source software programs for this application are now available (e.g., OJS, E-prints, OCS, DSpace, Fedora, etc.). There is a need to do research, distribution of tool kits, and advocacy in both Portuguese and Spanish. Implementation can be done by IBICT and partners, such as the IAP/ IANAS, UN GAID e-SDDC, DRIVER (in EU), Red CLARA. Potential funding might be obtained from the UN GAID regional office, European Commission, IAP Program, Moore Foundation (which supports the adoption of the Fedora software model), and others.

16. Accelerate scientific information cycles using common-use licenses, policies and social incentives to create an open access and use environment for research in Latin America.

This can be promoted by building on successful examples of institutions applying Creative Commons licenses (e.g., approximately 300 scholarly journals, the Uniprot database). The Science Commons is working with national research communities and open access publishers in the areas of science, technology, and medicine in developing countries.

17. Build capacity in intellectual property (IP) and open access (OA) by creating a partnership with stakeholders at the national level involving IP experts, ministry

officials, research funding organizations, universities, libraries, and OA publishers, together with international partners.

Intellectual property concepts are arcane. There is lack of knowledge about the role and scope of IP, what types of works are protected and in what contexts, and when common-use licenses are appropriate. The situation is complex and challenging, even for professionals in government and research. National research funders do not have open access in their mission or mandate. How to license for common use is still largely unknown. IP policy and legislation are controlled by commercial interests, mostly by OECD multinationals. Therefore, there is a need to promote partnerships for sound policy-making on IP and OA for public research.

This can be done by developing strategies and models to promote, communicate, and implement OA for publicly-funded research. Context-dependent rationales are the most effective approach, making the case from the self-interested perspectives of each community. Such a partnership should include coalitions at the national level, as described above, and at the international level (e.g., CODATA could coordinate with Information Age Publishing, Creative Commons/Science Commons, Access to Knowledge (A2K), Open Content Alliance, open access publishers, library organizations, other non-governmental organizations, and domain-specific programs and organizations). A working group could be established to develop such a partnership, perhaps on the Open Content Alliance model. The working group could hold an organizing meeting in Washington, DC and initiate a pilot project with Chile.

Session 4. Closing Presentations and Future Follow-up Activities

The closing meeting was chaired by Michael Clegg, Foreign Secretary of the National Academy of Sciences, USA. Speakers were Krishan Lal, President of CODATA[53] and Hernan Chaimovich, Vice President of the Brazilian Academy of Science[54].

Krishan Lal made a brief presentation of CODATA, ICSU's (International Council for Science) Committee on Data for Science and Technology. CODATA's mission is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. It is a non governmental organization and has 24 countries and 15 international scientific unions as members, 4 co-opted scientific organizations, and 13 supporting organizations from industry, government and academia. Paul Uhlir of the US NAS and Prof. Paul David of Stanford and Oxford Universities helped CODATA establish the new Global Information Commons for Science, initiative which focuses on legal, economic and institutional perspectives. CODATA also founded and runs the *Data Science Journal* a peer-reviewed electronic journal on data management and databases in science and technology.

Hernan Chaimovich stressed that ICSU's long-term vision is for a world where science is used for the benefit of all, excellence in science is valued and scientific knowledge is effectively linked to policy-making. In such a world, universal and equitable access to high quality scientific data and information is a reality and all countries have the scientific capacity to use these and to contribute to generating the new knowledge that is necessary to establish their own development pathways in a sustainable manner. Chaimovich believes that CODATA should develop a clear long-term strategy that focuses on key international data management and policy issues, giving special attention to the needs of developing countries. While developing its long-term strategy, a short-term CODATA focus on implementation of relevant aspects of Science in the Information Society Agenda for Action and follow-up to the World Summit on the Information Society II would be both appropriate and valuable.

---

[53] available at www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session12/CODATA.ppt

[54] available at http://www.cria.org.br/eventos/codata2007/apresentacao/Presentations_10-05/Session12/Brazilian%20Academy%20of%20Sciences.ppt

FINAL COMMENTS

The workshop brought together specialists from different disciplines of knowledge to demonstrate current practices and to discuss challenges to, and opportunities for, open and permanent access to data and information, focusing on *health*, *environment*, and *Latin America*. Many important initiatives were described, challenges and barriers, as well as opportunities, were discussed, and many action items were suggested.

One of the most positive aspects of the discussion was the number of initiatives that are already in place in Latin America and partner countries. One important asset is the necessary communication infrastructure that enables data access and exchange and distributed information systems through the Internet. Both CLARA Net - Cooperación Latino Americana de Redes Avanzadas (Latin American Cooperation of Advanced Networks) and RNP – Rede Nacional de Ensino e Pesquisa (National Network of Education and Research), responsible for the high-speed Internet connectivity for research and educational institutions in Latin America (CLARA Net) and Brazil (RNP), described their existing capabilities and future plans.

The cyber-infrastructure for continental collaboration is being further developed to include all countries and to create additional paths. It is already operational, and most countries are integrated. This is a tremendous asset. Integrating regional telecommunications networks and interconnecting National Academic Networks (NRENs) of the region with Internet2 in the US, and to the European Advanced Network (GÉANT), will accelerate the development of research capacity and the information society in Latin America. When facing the challenges of managing enormous volumes of data in a reliable and efficient manner, and allowing cross-institutional and cross- disciplinary collaborations, network infrastructure becomes essential. A key requirement is that this infrastructure be constantly supported, upgraded, and available as an essential support requirement for open and free sharing of and access to data and information.
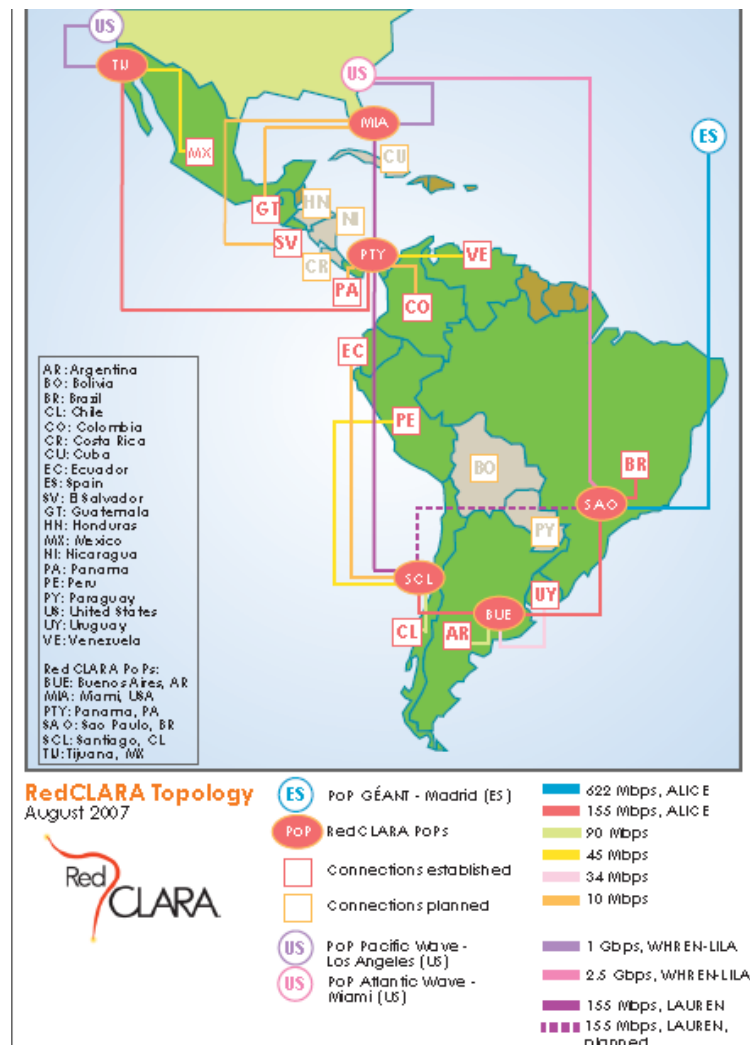
Figure 1. RedClara Topology – August 2007

The workshop discussions also focused on free and open access to journals. Collaborative approaches to open access, such as SciELO and Bioline International, were described as were some organizational and technical solutions, such as institutional repositories and virtual incubators of digital contents. Latin America does not have a strong editorial and publishing industry, and cannot afford access to the world's information. Open access approaches are needed, not only to provide access to the world's knowledge, but to make each country's research results available, both locally and globally.

Existing business models for research journals and data centers that restrict access to, and use of, publicly funded scientific information do not take advantage of the potential benefits of global open availability online, and consequently are slowing the progress of science and socioeconomic development. The question, therefore, is not whether open availability to such publications is better than access provided on terms that are economically, legally, and technologically restricted, but how effectively and quickly can open availability be institutionalized. This requires the restructuring of scientific communication in public research, which is gradually taking place. Openness should be the default rule, subject only to legitimate and well-justified exceptions.

Another problem concerning scientific journals is that scientific productivity indicators are primarily based on high-impact international journals. Publishing in national scientific journals is fundamental to communicate research results of national or regional interest and to stimulate the development and consolidation of research fields within a country. There is thus a need to include Latin American journals and those in other developing regions in citation indexing systems.

It is important as well to start implementing indexing methods with open access literature published under Creative Commons licenses, and evaluate the impact of such literature. Another suggestion

is to develop new value-added services, such as linking databases to open access literature, and implementing this in open, non-proprietary formats.

Thanks to the enormous progress on information and communication technologies one can also achieve open access to data collections themselves. Policies are needed for data that are born digital, to have mechanisms for digitizing legacy data, and for implementing long-term policies for the development and maintenance of data centers. A clear data policy at the governmental, institutional, and personal and disciplinary levels is required. In addition, procedures for preparing data and information for permanent access need to be merged with research planning.

Government-wide policies that are focused on ensuring that the value of publicly funded research is maximized have begun to emerge worldwide. These policies are exploring the possibilities for advancing the conduct of research by providing stable, long-term archives for research outputs, and in enhancing access to, and use of, research results. Making data and information freely and openly available requires planning and adequate resources, expertise and infrastructure, and long-term stable financial support for collections and information systems. Project based support, by itself, will not be sufficient to ensure long-term availability.

Open data systems are normally developed to fulfill common objectives outlined by global and regional initiatives. There are also needs to fulfill national demands related to health, environmental, and biodiversity information and data. In order to meet these challenging demands, local stakeholders need to understand and use such data in decision-support systems and policy-making processes. Effective data and information organization and dissemination is essential for their optimal application.

The workshop discussions also acknowledged there is a general lack of knowledge at the working researcher level about intellectual property, licenses, and contracts. Therefore, in addition to long-term data preservation policies, clear rules for sharing data with consistency and understanding of intellectual property rights need to be established.

A reward system and various incentives for stimulating open publishing and sharing of databases need to be established as well. Valid metrics should be developed to help understand the impact of access to data and information. CODATA might consider establishing a Task Group on this topic.

Health and biomedicine are fields that require information and communication technology (ICT) to help integrate both the institutions and the data. At the workshop, PAHO presented the use of public health indicators for decision-making that depend on access to local data and to data from other countries. Examples of scientific literature networks and telemedicine were also demonstrated. This is a growing field that depends largely on ICTs and requires capacity building throughout the region.

Environmental and geospatial data perhaps present the greatest technical challenge to ICT infrastructure. The amount of data is massive. These data need to be integrated with other knowledge sources to promote progress in many fields. There are numerous challenges and barriers to providing open and permanent access, and some ideas for making progress were suggested.

Finally, open access to biodiversity data and metadata is advancing in the region. At the global level, the Global Biodiversity Information Facility (GBIF) is integrating species and specimen data worldwide through a partnership with a number of GBIF nodes in the world. There are 5 GBIF nodes in Latin America (Argentina, Colombia, Costa Rica, Peru, Mexico) and 5 information centers that are already active in integrating biodiversity data (Brazil, Colombia, Costa Rica, Peru, Mexico) following accepted international standards and protocols. Research institutions located in Amazonian countries, along with institutions with relevant specimen and data holdings from the region, are working on the development of an Amazon Basin Biodiversity Information Facility.

Together, all of the suggestions by the experts at the workshop for promoting open and permanent access to scientific information in Latin America were intended to focus more attention and foster greater cooperation on all of these issues in the region.

ANNEX: ACRONYMS USED

ABBIF - Amazon Basin Biodiversity Information Facility

ACTO - Amazon Cooperation Treaty Organization

BIREME – Biblioteca Virtual em Saúde (Brazil)

CBERS - Chinese-Brazilian Earth Resources Satellite Program

CEDIA - Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado

CGEE - Centro de Gestão e Estudos Estratégicos

CI – Conservation International

CLARA - Cooperación Latino Americana de Redes Avanzadas (Latin American Cooperation of Advanced Networks)

CODATA – Committee on Data for Science and Technology

CONICYT – Comisión Nacional de Investigación en Ciencia y Tecnología

CPTEC – Centro de Previsão do Tempo e Estudos Climáticos, Brazil

CRIA – Centro de Referência em Informação Ambiental, Brazil

e-SDDC - Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries

EU – European Union

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo (The State of São Paulo Research Foundation)

Fiocruz – Fundação Instituto Oswaldo Cruz

FUNDATEL – Fundación de Telemedicina

GBIF - Global Biodiversity Information Facility

GÉANT - European Advanced Network

GEOMA – Rede Temática de Pesquisa em Modelagem Ambiental da Amazônia

GOES - Geostationary Operational Environmental Satellite

GEOSS - Global Earth Observation System of Systems

GIS - Geographic information systems

GSDI – Global Spatial Data Infrastructure Association

IABIN - Inter-American Biodiversity Information Network

IAI - Inter-American Institute for Global Change Research

IBGE – Instituto Brasileiro de Geografia e Estatística

IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia

ICSU - International Council for Science

ICT – Information and Communication Technology

IUCN - The World Conservation Union

IDB – International Data Base

IDRC - International Development Research Centre

IIA - Instituto de Informática e Automação, Brazil

IIAP - Instituto de Investigaciones de la Amazonía Peruana

ILOAR - Ibero-Latin American Open Access Repositories

INPA – Instituto Nacional de Pesquisas da Amazônia

INPE – Instituto Nacional de Pesquisas Espaciais, Brazil's National Institute for Space Research

ISI – Information Sciences Institute

Manis - Mammal Networked Information System

MODIS - Moderate Resolution Imaging Spectroradiometer

MSG – Meteosat Second Generation Satellite

NDVI - Normalized Difference Vegetation Index

NIH - National Institutes of Health, USA

NOAA – National Oceanic and Atmospheric Administration, USA

NPOESS - National Polar-orbiting Operational Environmental Satellite System

NRENs - National Academic Networks

NSF – National Science Foundation, USA

OAS - Organization of American States

OECD - Organization for Economic Co-operation and Development

PAHO - Pan American Health Organization

PAIGH Pan-American Institute for Geography and History

PPBio –Programa de Pesquisa em Biodiversidade (Program for Planned Biodiversity Studies)

RAINFOR/ATDN- Red Amazonica de Inventarios Forestales/ Amazon Tropical Diversity Network

RANPA - Amazon Protected Areas Network

Red CLARA - Latin American Network for Science, Education and Innovation

RENATA - Red Nacional Académica de Alta Tecnologia en Colombia

RIPSA - Rede Interagencial de Informações para a Saúde (Scientific Literature for the Interagency
Network of Health Information)

RNP – Rede Nacional de Educação e Pesquisa (Brazil's National Education and Research Network

RUTE - Rede Universitária de Telemedicina (Brazil)

SciELO – Scientific Electronic Library Online

SIAMAZONIA - Peruvian Amazonian Biodiversity Information System

SiB - Biodiversity Information System of Colombia

SIGMA (Sistema de Informação Geográfica Aplicado ao Meio Ambiente - Geographic Information System
Applied to the Environment)

SPARC - Scholarly Publishing and Academic Resources Coalition

SRTM - Shuttle Radar Topography Mission

TIDIA – Tecnologia da Informação para o Desenvolvimento da Internet Avançada

UN – United Nations

UN GAID - UN Global Alliance for ICT and Development

UNAMAZ - Association of Amazonian Universities

UNDP - United Nations Development Programme

ViNCES - Virtual Network Center of Ecosystem Services

WB – World Bank