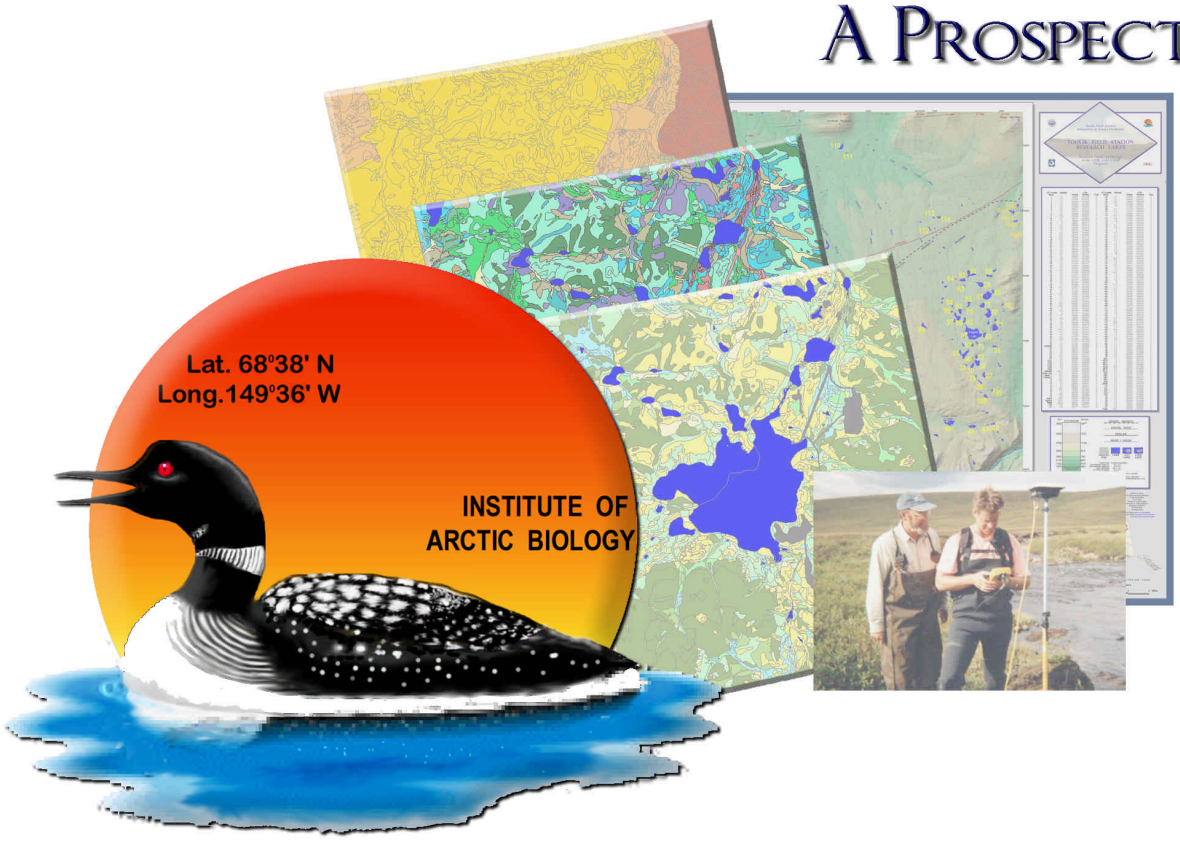
Alaska Center for EcoInformatics

- ACE -

A Prospectus



Toolik Field Station GIS

Andrew W. Balser, GIS & Remote Sensing Manager
Institute of Arctic Biology
University of Alaska Fairbanks
May 5, 2006

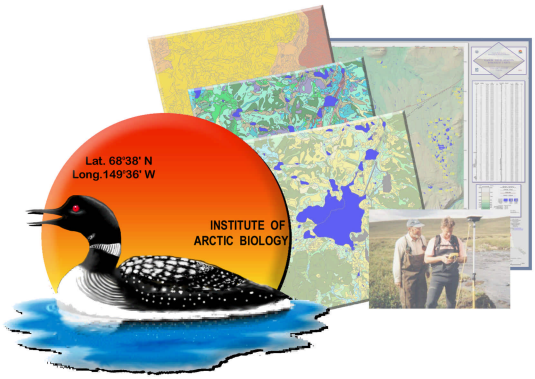
Alaska Center for EcoInformatics

- ACE -

A Prospectus

University of Alaska

2006



Andrew W. Balser
GIS & Remote Sensing Manager
Toolik Field Station

Institute of Arctic Biology, Campus Box 757000
University of Alaska Fairbanks
Fairbanks, AK 99775-7000

Andrew.Balser@uaf.edu
(907) 474-2466 (phone)
(907) 474-5513 (fax)
http://www.uaf.edu/toolik/gis/

Contribution & Comments:
Dr. Brian M. Barnes, Director, Institute of Arctic Biology, UAF
Dr. Donald A. 'Skip' Walker, Director, Alaska GeoBotany Center, UAF
Dr. Falk Huettmann, Institute of Arctic Biology, UAF

CONTENTS

1.0 EXECUTIVE SUMMARY

OBJECTIVE:

Maintain and expand the University of Alaska's leadership role in multi-disciplinary ecology, while linking ecological research with broader natural science research

PROPOSAL:

We propose the Alaska Center for EcoInformatics (ACE) at the University of Alaska (UA). ACE will be broad-scope, incrementally built and will rely heavily on community feedback, pre-existing strengths, and natural synergies among campuses and programs.

BENEFITS:

Increased federal funding for research associated with broad-based initiatives
(e.g. IPY, NSSI, LTER, CUAHSI)

More analytical results and publications per research dollar

Increased effectiveness and efficiency for multiple research programs

Support to help retain good faculty, staff and students

Better and more sophisticated collaborative science

Eventual linkages with BioMedical and Geophysical research
(via BioInformatics, GeoInformatics)

TIMING:

EcoInformatics in the U.S. and abroad is developing very rapidly. Both the need and the capacity for it are now great enough to make implementation plausible and urgent.

IMPLEMENTATION:

Build incrementally, leveraging pre-existing strengths and addressing the most fundamental needs first

Staff ACE initially with four complementary personnel, allowing enough staff for program stability while keeping it small enough to remain manageable and responsive.

Foster constant communication and interaction within and beyond UA programs – use the skills and experience of other institutions whenever possible to improve the program, and increase its interoperability with similar programs at other institutions.

2.0 BACKGROUND

EcoInformatics is the handling of knowledge and information to create seamless integration and transition among disparate research disciplines, through different temporal scales, and among distant research nodes. As a robust fusion of personnel, data and technology within a cogent conceptual framework, EcoInformatics expands and advances ecological, biological and natural sciences, and natural resource management within and beyond a given home institution. Services include: 1) multi-disciplinary, multi-temporal, multi-locational synthesis, 2) active advising and consultation to project-based research, 3) scripting and tool development, and 4) a stable legacy platform for data, knowledge and services. Benefits include: 1) new scientific discovery, 2) identification of collaborative opportunities, 3) increased efficiency and effectiveness of constituent research, 4) improved outreach, and 5) expanding the relevance of constituent project results.

Numerous initiatives already underway are critical enablers to EcoInformatics in Alaska and the Circumarctic. Web portal development, Spatial Data Infrastructure (SDI) enhancement, metadata tool refinement and related efforts all serve to make existing data easier to find and access. EcoInformatics leverages these services to develop direct linkages among disparate datasets, enabling direct ecological analysis using multi-disciplinary data. For example, a web portal may serve a number of different datasets containing nitrate measurements. These numbers may come from aquatic, pedologic and vegetation studies, and comprise different timescales of measurement. As such, they are standalone datasets with no direct relationships, and are difficult to use together. EcoInformatics uses scientific expertise to identify logical, process-based linkages among these datasets so the data may be translated into a format where information is equivalent and comparable. Technological expertise then builds tools for researchers to conduct specific analyses and syntheses of these data. EcoInformatics also includes a personnel infrastructure to help researchers, administrators and stakeholders use the existing knowledge base and tools.

EcoInformatics provides a critical complement to project-based research by meeting common needs through a continuing, stable program. Synthesis, analysis, query and comparison, programming and scripting, and advisory functions are provided through a service center available to individual investigators, labs, institutes and agencies. Previously non-existent services are provided, while constituents are relieved of supplying their own expertise in a redundant fashion. As a program, it is also a natural communication node for similar initiatives in other US and international institutions.

The University of Alaska will need a robust and strongly supported EcoInformatics program to maintain active participation and leadership in Observing Networks, the Long Term Ecological Research (LTER) program, the Consortium of Universities for Advancement of Hydrologic Science (CUAHSI), and Spatial Data Infrastructure (SDI) initiatives. The coming International Polar Year (IPY, 2007 - 2008) offers particular incentive for an EcoInformatics program now to proactively coordinate data, tools and personnel services for inter-linked IPY projects. Expansive by nature, IPY projects would be natural beneficiaries of EcoInformatics. Smoother, efficient collaboration and data sharing and overall coordination of Alaska based IPY projects is a clear benefit to researchers, the University and the state of Alaska.

EcoInformatics is a new field actively defining its structure and finding its boundaries. Initiatives range from small, single-institution working groups to large, international, multi-agency consortia developing tools and standards to drive policy decisions. Regardless of scope, any effective EcoInformatics program recognizes that redefining structure and shifting boundaries are not just temporary responses to early growth, but rather the healthy perpetual state of a service-oriented program that thrives on adaptive development. One size does not fit all, and each EcoInformatics implementation needs and deserves to be specifically tailored to the distributed community it serves while developing interoperability with other informatics programs. It is becoming axiomatic that initiatives leveraging technology for science are most successful when scoped copiously and built incrementally, gathering continual community feedback[1,2]. ***We propose a thoroughly-scoped, incremental approach at the University of Alaska, relying heavily on community feedback, strides already made, and on natural synergies among existing programs***. The outcome will involve available services at the University of Alaska statewide level, and the program will maintain close ties with state, regional, national and international initiatives involving EcoInformatics, IPY and Observing Networks, and natural resources management.

Many components necessary for a robust UA EcoInformatics program already exist. This program will be a direct complement to the Geographic Information Network of Alaska (GINA), and the program will actively tap the talents and institutional knowledge of schools and departments on each campus. The role of the EcoInformatics program will be to coordinate new and existing data and technology with a stable platform of personnel to provide targeted tools, knowledge integration and advisory services germane to ecological sciences now, while preparing for an inevitable future merger with other 'informatics' areas (i.e. BioInformatics, GeoInformatics).

3.0 ALASKA CENTER FOR ECOINFORMATICS (ACE)

3.1 Vision

Several key characteristics will drive ACE's development and success. The primary goal is functional linking of knowledge, data and personnel among labs, programs, departments and campuses in the UA system. Also critical is recognition that the eventual sphere of influence extends far beyond ecology into other natural sciences, natural resources management, and policy. The approach is a flexible, interoperable, service-based industry; explicitly making it easier for researchers, managers and the general public to learn interactively using broad arrays of available data and knowledge. This is an important distinction. Comprehensive data discovery, display and distribution services are already in development at the UA level (e.g. web portals). Aggregating myriad datasets in a searchable, documented environment like a web-portal is a non-negotiable prerequisite for meaningful future science and management. ***EcoInformatics fills a critical, different, complementary role predicated on access to web-portal resources.***

Research science is typically conducted on a project by project basis. While synthesis may be a key element in many projects, the nature of the system does not put the highest priority on linking each project to a larger, stable framework of data, technology and knowledge. Rather, the highest priorities are to fund, conduct, complete, and publish the project. This system makes perfect sense given the pathway to funding acquisition, and it produces fantastic results which are clearly conveyed through scientific literature. The downside is that there is less incentive for investigators, graduate students, and technicians to invest personal overhead in fully exploring the broader picture of data and knowledge related to their research. New technologies, new tools, and unfamiliar datasets take an enormous amount of time to assimilate; time that is difficult or impossible to find. Results of this include:

1) Many projects lack integrative analyses that could vastly expand the relevance of the results.
2) Many projects include some integration and synthesis, but far less than is actually possible.
3) Many datasets do not make it to a publicly accessible location once the project ends.
4) Critical collaborative opportunities and scientific discoveries are missed.
5) Much potential of existing science and the funding that supports it goes unrealized.

EcoInformatics helps bridge project level gaps by providing program level services. This explicitly includes personnel infrastructure to help users maximize the potential of the data and knowledge that ecology provides.

3.2 Activities & Implementation

ACE will be built incrementally. Initial focus will be on simple, fundamental services which will a) provide immediate function and benefit, b) expose as many people as possible with the program (fostering involvement), and c) lay groundwork for more sophisticated services. Specific tasks will largely be defined by members of the community as program scoping progresses. Early services should generally involve short turn-around, and a tendency to "under-commit and over-achieve". This approach has been very effective at growing community involvement, interest and broad-based support for program level services (e.g. Toolik Field Station GIS & Remote Sensing Program). As the

program takes shape and gains a track record, ACE will become more integrated with entire projects, with involvement beginning at the proposal stage and continuing through analysis, interpretation and data delivery / archiving. Ultimately, ACE will serve as a key additional resource to help coordinate interdisciplinary research and identify new opprtunities.

Some examples of services/functions include:

1) Help researchers write and modify analysis scripts that leverage cross-disciplinary datasets
2) Help researchers identify appropriate datasets from outside their own field, provide context, tools and consultation for incorporating them into their own analyses
2) Amass a library of scripts and applications – help researchers identify and use them
3) Provide leadership developing web-based resources (e.g. ToolikInfoPortal 'TIP', w/ GINA)
4) Scope and define logical linkages among disparate ecological datasets (e.g. appropriate key fields, measurement units, temporal increments, summary statistics, database integration)
5) Foster connections among web-based data delivery and analysis resources (i.e. develop appropriate seamless linkages among web portals and distributed-node data infrastructures)
6) Develop tools and framework to expand current research in both space (e.g. landscape extrapolation) and time (via historic datasets and predictive models)

Future responsiveness and flexibility is a central requirement for the long-term success and continued value of the program. We recognize from the outset that scoping this program must involve careful vision to prepare data and services to be most easily adapted to unforeseen uses. Technical solutions to data management and distribution must be robust enough to fit into multiple applications easily, rather than committing to an elaborate software and database system that will become obsolete and unwieldy.

3.3 Structure

We believe that appropriate staffing is instrumental to ACE's success. The key is to provide a range of critical skill sets within a core group. The group should be small enough to function as an efficient team, while large enough to cover all primary functions and provide institutional stability. Constant communication with constituents will be a central requirement for every staff member. As a service-oriented program, continual and proactive outreach to pertinent parties is prerequisite to successful implementation, continued flexibility and future growth.

Broad perspective and guidance will come through regular interaction with the ACE Steering Committee, to be comprised of investigators, technical personnel and science administrators. Bi-monthly meetings will keep the program apprised of upcoming changes within UA and in ecological science in general, and will assure the Steering Committee that implementation and services are continuing in the right direction.

3.3.a Staff

We propose a staff of four full-time employees, hired over a two year period. This is an ideal number for two reasons. ***First, four staff will lend stability to the program; if an individual leaves, there is sufficient institutional knowledge to bridge the gap. Second, four is a small enough number to be nimble and manageable, effective but not top-heavy.*** These four will

provide a suite of EcoInformatic services to directly help researchers, administrators, students and the public. Personnel expertise will be complementary to GINA, with whom we will frequently interact. Skill overlap will be enough to promote a smooth working relationship without unnecessary replication. Moreover, we will actively recruit staff with dual backgrounds in both natural science and technology. ***It is critical for all staff to be fluent in both a technical field and an ecological/biological discipline. The common criterion for all staff is to interact with researchers frequently and effectively, translating technological application to ecological knowledge.***

3.3.a.1 Director

The Director will be the primary conduit connecting ACE staff with its Steering Committee, Institute and Department level leadership, GINA administration, and agency personnel. The Director will translate the goals developed in consultation with the Steering Committee into a day to day action plan, orchestrating task and project level activities into a cohesive and responsive program. The Director will also oversee all initial project development between the client base and ACE staff, ensuring that final products will be scientifically sound, technologically effective, and fit the client's needs. ***The Director's background should include exposure to multiple ecological disciplines to maximize the position's effectiveness***.

The Director will submit a semi-annual report to IAB, GINA, and UA administration as well as to the ACE Steering Committee. Aside from simply reporting activities, initiatives and expenditures, the role of this report will be to help top-level personnel better identify big-picture opportunities among university programs and research ventures.

The Director, therefore, will possess a) technical skills, b) a breadth of ecological/natural science knowledge, and c) developed capacity for public communication and program development. ***While the director will engage in technical activities, the key role is interface among research, administration and the public.***

3.3.a.2 Core Staff

The remaining three full-time positions will be filled on the basis of complementary skills and background. It is expected that each staff member will have both technical skills and solid foundation in an ecological field. The final combination of four staff should include the following skills and discipline backgrounds collectively. ***A very positive observation is that these skills and backgrounds already exist within the UA system, allowing us to tap the expertise we already have.***

Skills

- Remote Sensing and GIS
- Scripting for statistical / spatial analysis (e.g. SAS, Python - ArcGIS, IDL)
- Database administration (e.g. Oracle, Postgres)
- Web design / Java
- Internet Mapping (e.g. MapServer, IMS)

Disciplines

Plant or Terrestrial Ecology
Aquatic Ecology
Population Ecology
Landscape Ecology
Systematics

3.3.b Steering Committee

The Steering Committee will be a critical communication conduit between the staff, the research community, and administration. Initially, it should be comprised of up to eight people including at least one representative from each campus (UAF, UAA, UAS), and at least one UA statewide science administrator. The combined expertise should include most of the above-listed skills and disciplines, drawn ideally from people holding directorships and leadership positions across institutes and departments.

Primary functions for the Steering Committee are to set long-term goals, make strategic-level development recommendations and provide richer context for determining the ultimate role of the program within the UA system and beyond.

The committee would presumably meet bi-monthly in the first year, scaling back to an annual meeting once ACE has developed a clear shape and a track record. Explicit goals within the first year should include bringing outside people to Alaska to interact with the committee to forestall mistakes and maximize interoperability with other institutions and agencies.

3.4 Collaborators

Individual collaborators will eventually include a large percentage of those involved in ecological, biological, natural resource and environmental research in the UA system. It is also expected that agency personnel within Alaska and researchers from outside institutions who work here will develop active relationships with ACE. The following institutions are either obligate or likely candidates for collaboration.

UA System
Institute of Arctic Biology (IAB), UAF
Geographic Information Network of Alaska (GINA), UAF
Environment & Natural Resources Institute (ENRI), UAA
Environmental Sciences Program, UAS
Institute of Northern Engineering (INE), UAF
School of Fisheries and Ocean Sciences (SFOS), UAF
Water and Environmental Research Center (WERC), UAF
UA Geography Program
School of Natural Resources and Agricultural Sciences (SNRAS), UAF

Alaska Agencies
Alaska Dept. of Fish & Game (ADF&G)

Alaska Dept. of Natural Resources (ADNR)
Alaska Dept. of Environmental Conservations (ADEC)

Federal Agencies
US Fish & Wildlife Service (FWS)
US Geological Survey, Biological Resource Division) (USGS-BRD)
US National Park Service (NPS)
US Forest Service (USFS)
US Environmental Protection Agency (EPA)

4.0 ECOINFORMATICS BACKGROUND

Ecology is the pursuit of a process-based understanding of the natural world. Theoretically, the need for EcoInformatics has existed as long as the field of Ecology, but in practical terms the seamless integration and transition across disciplines, through temporal scales, and among spatial locations has only recently become possible. Improved computing and networking, automated data collection, and development of ecological principles offer opportunities for researchers to simultaneously expand knowledge within a specific area and for a broad suite of disciplines. Grasping the realm of possibilities, scoping effective programs, and succeeding in implementation are all challenging tasks, yet the benefits of a robust EcoInformatics program dwarf the initial investment through increased knowledge, capacity and efficiency.

4.1 Brief History

A probable starting point for EcoInformatics is the Rio Convention of 1992. Researchers interested in biodiversity identified the need for a comprehensive, robust platform for tracking changes in species population and composition on a worldwide basis. This requires, at minimum, a large body of data, exhaustive organization, and open collaboration. The Global Biodiversity Information Facility (GBIF) is a result of this recognition, and provides data, services, news and interpretation through its web portal (http://www.gbif.org/). A key feature of GBIF is its successful incorporation of information from a wide variety of original sources. This 'distributed node' architecture is generally considered the most effective approach to data synthesis and integration. We have been associated with numerous panels and working groups since 2001 that have all identified distributed nodes as a key characteristic of successful efforts in Alaska and the arctic. Many of these comments are well summarized in the white paper led by Mark Sorenson of the University of Redlands, and to which we were contributors (http://www.uaf.edu/toolik/gis/CircumArcticGIS_2004-06-13.pdf).

Since 2000, EcoInformatics have been implemented in a variety of ways at a multitude of scales. Each successful effort addresses the needs of initial stakeholders, handles the pre-existing reality of data format and availability, lays the groundwork for future interoperability with other Informatics programs, and remains flexible to necessary future adaptation. Initiatives begin predominantly within academic institutions, though their benefits extend outward to agencies, K-12 education and the public. The following examples illustrate the variable scope and scale, and underscore the point that interoperability and unique, individual implementations are not mutually exclusive.

4.2 Case Studies

4.2.a EcoInformatics.org

EcoInformatics.org is a collaborative effort sponsored by the Long Term Ecological Research (LTER) network and the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California Santa Barbara (UCSB) (www.ecoinformatics.org). Housed at NCEAS, a premier ecological think-tank, this group focuses on high-end application and model development in support of sophisticated analytical synthesis. They produce tailored software applications, facilitate metadata creation and update, and provide a framework for better data discovery and integration. An important observation is that they are focusing on very high-end services which presume a high level of data organization, documentation and completeness. Most owners of scientific data – research institutions and agencies alike – do not yet have this level of preparation. UA and Alaska in general are in this position, and will need an EcoInformatics program to provide leadership and assistance in this regard.

Examples of primary projects include:

1) The Knowledge Network for BioComplexity - http://knb.ecoinformatics.org/index.jsp (data network)
2) The Kepler Project - http://kepler-project.org/ (data integration/modeling software)
3) Jalama - http://jalama.ecoinformatics.org/ (metadata creation/update – germane to ecology)
4) Science Environment for Ecological Knowledge - http://seek.ecoinformatics.org (cyberinfrastructure)

There are several key characteristics of this initiative illustrating a very savvy implementation of EcoInformatics. Most importantly, their central goal involves functional linking among disparate disciplines within ecology, and recognition that the eventual sphere of influence extends well beyond ecology into the other natural sciences, natural resources management, and policy. The approach is as a service-based industry; they are explicitly trying to make it easier for researchers, managers and ultimately the general public to learn interactively using broad arrays of data and knowledge. This is an important distinction. Aggregating myriad datasets in a searchable, documented environment like a web-portal is a fundamental prerequisite for meaningful future science. EcoInformatics fills a critical, different, complementary role predicated on current or future access to such resources.

4.2.b The EcoInformatics Collaboratory

The EcoInformatics Collaboratory at the University of Vermont's Gund Institute for Ecological Economics (http://ecoinformatics.uvm.edu/) addresses the need for process-based environmental knowledge to understand human interactions and impacts on the natural world. With an eye on driving policy decisions, this collaboratory provides similar functions to EcoInofrmatics.org (synthesis of disparate disciplines, tools) but from a different starting point. Under the philosophy of using rich, independent, comprehensive data for responsible public decision-making, they provide the necessary tools and information, with consultation on request. A particular appeal to this implementation is that the validity of the information and knowledge isn't compromised by its source. Since the services are entirely independent of actual decision making agencies, the results can be seen as impartial.

Examples of primary projects include:

1) Economic Valuation of Ecosystem Services
http://ecoinformatics.uvm.edu/projects.html (functional data synthesis with analysis tools)
2) Integrated modelling of Earth's social, economic and ecological dynamics
http://ecoinformatics.uvm.edu/projects.html
3) Environmental and Agricultural modelling in the European Union
http://ecoinformatics.uvm.edu/projects.html

4.2.c The EcoInformatics Initiative

The EcoInformatics Initiative (http://ecoinfo.eionet.europa.eu/) is a very broad, top down effort to 1) enhance interoperability among EcoInformatics programs, and 2) bring public-sector decision makers together at an international (US and EU) level to establish protocols and encourage use of synthesized ecological information within a multi-agency, international context. Their activities revolve less around providing specific services to user groups, but instead cover big-picture questions relating to valid information use and likely needs at a global scale. As a top-level scoping effort, they are in a position to provide guidance to most or all EcoInformatics initiatives as they attempt to attain interoperability with other regional, national and global efforts.

4.2.d International Assn of Vegetation Science (IAVS): EcoInformatics Working Group

In direct contrast to the EcoInformatics Initiative, the IAVS EcoInformatics Working Group (http://www.iavs.org/workg.htm#Ecoinformatics) is a small-scale, targeted effort to help scientists in a specific field integrate their research with EcoInformatics. Their efforts include some limited tool development and recommendations for standards, but they are primarily spreading awareness among their constituency, and filling the consultant role in actively assisting their colleagues in getting involved.

4.3 Summary

Together, these examples illustrate 1) the variability of EcoInformatics implementation, 2) the value in tailoring solid principles to specific needs and communities, 3) multiple pathways to effective interoperability, and 4) EcoInformatics is, and should remain, a dynamic, responsive and evolving field. The Alaska Center for EcoInformatics will build upon successes of efforts like these, while crafting a program suitable for our specific and changing needs here in Alaska.

5.0 LINKS

5.1 EcoInformatics

EcoInformatics.org	http://www.ecoinformatics.org
The EcoInformatics Collaboratory	http://ecoinformatics.uvm.edu/
The EcoInformatics Initiative	http://ecoinfo.eionet.europa.eu/
IAVS EcoInformatics Working Group	http://www.iavs.org/workg.htm#Ecoinformatics

5.2 UA and Alaska

Toolik GIS	http://www.uaf.edu/toolik/gis/
Institute of Arctic Biology	http://www.iab.uaf.edu/about.php
GINA	http://www.gina.alaska.edu/
ASDI IPY Proposal	http://www.ipy.org/development/eoi/details_print.php?id=265
UA Statewide Geography	http://www.geographyua.org
School of Nat. Res. & Ag. Sciences	http://www.uaf.edu/salrm/

6.0 REFERENCES

1. *Recommendations for a Geographic Information Infrastructure to Support Arctic Research: Outcomes of the Arctic GIS Workshop, 22 – 24 January 2001, Seattle, WA.* Mark Sorenson, William F. Manley, Renee D. Crain and Wendy K. Warnick (eds.). The Arctic Research Consortium of the U.S. (ARCUS). Fairbanks, AK. 2001. 49pp.

2. *Arctic Spatial Data Infrastructure (SDI) for Scientific Research:* Mark Sorenson, Sabine Hoover primary authors. The Arctic Research Consortium of the U.S. (ARCUS). Fairbanks, AK. 2004. 41pp.