

# 30 Information Technology and the Ten Grand Research Challenges for the 21<sup>st</sup> Century

**Leonard Krishtalka**

Doubling the budget of the science agencies is a terrific goal. But first, we should consider for what purpose we are doubling that budget. For the National Science Foundation and other nondefense and non-biomedical agencies, here is my answer: I propose we identify ten grand challenges in research for the 21<sup>st</sup> century. And, more importantly, we should also identify what they have in common.

## The Ten Grand Challenges

Let us will begin with understanding and modeling the physical universe. There are two grand challenges here:

1. What is the origin, structure, and fate of the universe?
2. What is the fundamental structure of energy and matter?

The next five grand challenges address understanding and modeling Earth's environment.

3. We need to decipher Earth's physical systems—its climates, geology, hydrology, ocean systems, and so forth. How do they work? What are their causes and processes?

4. What is the diversity of life on Earth? What if an alien landed on Earth, in the middle of Kansas, looked around at the beautiful tall grass

---

*Leonard Krishtalka is director of the Natural History Museum and Biodiversity Research Center at the University of Kansas where he is also professor of ecology and evolutionary biology. This article is based on remarks delivered at the 26<sup>th</sup> Annual AAAS Colloquium on Science and Technology Policy, held May 3–4, 2001, in Washington, DC.*

prairie and the Flint Hills, saw the lush wildlife, and asked: “What are all the kinds of life on Earth?” We would have to answer, “We don’t know.”

After 300 years of the biological exploration of this planet, we still do not know most of the life on our planet. We have described about two million species on Earth, but we know that Earth has between 15 and 50 million species of plants, animals, and microbes. So we are now dealing with approximately five to ten percent of the evidence in biology and biodiversity science. How far would chemistry have gotten with five to ten percent of the Periodic Table of the Elements (nothing after boron)?

Recently, we have witnessed a good deal of excitement about the possibility of life on Mars. This is terrific science. But we have an explosion of life under our feet here, on this planet, that we are ignorant about. Isn’t it time for a census of the life on this planet—all of the species of plants, animals, and microbes that inhabit Earth? Mars and its alleged relics of life will still be here in 50 years. But, according to most biodiversity scientists, at the current rate of species extinction on Earth, we have about 50 years before much of a census of life on Earth becomes a paleontological exercise.

5. We need to understand the tree of life. The tree of life makes sense of all of biology. It tells us the evolutionary kinship of every single species of plant, animal, and microbe. Most importantly, the tree of life turns the descriptive sciences of biology, biomedicine, biophysics, and biochemistry into predictive sciences. Every phylogeny, every branching pattern on the tree of life, is a prediction of all the properties that those organisms possess, from genomics to ecosystem science.

6. The sixth grand challenge is to decipher the language of life. We must go beyond human genomics to organismal genomics, and beyond the genome to functional genomics, developmental pathways, physiologies, and so on.

7. The final grand challenge in modeling and understanding the environment is: How do these four phenomena—Earth’s physical systems, diversity of life, tree of life, and language of life—interact to form the planet’s web of life? What is the interactive complexity between physical systems on Earth, biodiversity systems, genomic systems, and ecosystems? How does this interactive complexity shape the global environmental systems on which all of life depends? If medical science is concerned with the health of one species on Earth, biosphere science is concerned with the evolutionary and ecological pulse of Earth’s other 15–50 million species.

8. The eighth grand challenge focuses on the one species on Earth that has had the greatest impact on the planet—*Homo sapiens*. From urban systems to agrosystems, we need to understand and model human systems, or human ecology.

9. The ninth grand challenge is in neuroscience. We need to understand the brain, and not just the human brain, but brain structure and function across organisms. We need to learn what neuroscience can tell us about designing the next generation of artificial thinking machines. (This will be addressed in more detail below.)

10. Finally, if we are going to tackle these grand challenges, the 10<sup>th</sup> and greatest grand challenge is understanding and modeling complexity. How do we integrate Earth's systems and human systems into a predictive modeling regime for the environmental management of the planet? Essentially, this is a synthetic challenge to understand and predict how the planet's physical and biological diversity and human ecology shape the global environmental systems on which all of life depends. This knowledge is critical to science and society and has never been more important to humans than it is today. We need this integration if we are to sustain natural environments and resources, improve human health, ensure economic and political stability, and enhance the quality of human life. Urgent need for this knowledge increases daily as the conversion of natural systems to human-managed systems accelerates the decline of biological diversity and causes large-scale alterations of Earth's biosphere.

Of course, every list of ten grand challenges should have an 11<sup>th</sup>. The 11<sup>th</sup> grand challenge involves informing policy and the broadest dissemination of this knowledge. We need the results of the ten grand challenges to inform effective planetary management. We also need a way to provide a knowledge server for the planet, giving universal access to planetary information. We already do it for the weather. We need a "Weather Channel" for Earth and human systems. We must provide this information to every citizen of this world—the ultimate democratization of knowledge. (This will be discussed in more detail later.)

What do these grand challenges have in common? Information technology. As students of anthropology, we were once taught that the hallmark of human evolution was language, or making fire, or making tools. When we saw that chimpanzees could make tools, we thought that the hallmark of human evolution was making tools to make tools. But all of those human characteristics are symptoms, not hallmarks. The hallmark of human evolution has been information management, from

hominid beginnings four million years ago to today. We have managed data, information, and accumulated knowledge better than any other species on Earth.

These ten grand challenges will produce an exponential increase in data and information across disciplinary domains. They will challenge us to provide access to that data and information, because ultimately the access to the information is as important or more important than the information itself. It is human evolution continued. If we cannot have access to the information for research and education, then we may as well have a locked library. These grand challenges will challenge us to integrate that data and information across knowledge domains. They will challenge us to apply analytical tools to model complex systems and simulate complex phenomena that were hitherto intractable. In this way, we will be turning what has been predominately a descriptive scientific enterprise into a predictive one. Information management—informatics—is the chassis on which that can occur.

How will this affect people? We are seeing and will increasingly see a growth industry in scientists who share rather than hoard data across disciplines, who build and use analytical modeling tools across knowledge disciplines, and who work in collaborative, cross-domain groups. In a sense, we are seeing science not imitate art, but science imitate its new medium, the informatics medium. We are seeing science expand from the single-discipline scientist and his/her single-fiber desktop connection to the collaborative research group networked on the Internet.

Informatics is revolutionizing our notion of time in all aspects of research on the ten grand challenges. We will increasingly see and are already witnessing real-time data access, integration, and analysis on the Web—not in months or days, but in minutes and seconds. We have real-time modeling and effects-prediction over the Internet. We also have real-time dissemination of research results. They can be tested by the research community in real time for immediate feedback to the researcher or team. This can, of course, enable real-time policy discussions and decisions based on real-time science.

Finally, a revolution in supercomputing will be instrumental in helping us meet the ten grand challenges, because the next generation of supercomputers will be the first generation of the new thinking machines of the future. In 2001, the best supercomputers are capable of three trillion operations per second. In 2010, they will be capable of 1,000 trillion operations per second according to scientists at Los Alamos National Laboratory. These supercomputers will not just be better cal-

culators, as has been their evolution until now; they will be new kinds of thinking machines. They are going to provide, for example, the capability of large-scale simulation. This could be a living object pieced together at atomic scale; a higher animal's brain assembled by every neuron and synapse; or the biosphere and its composition, structure, and processes simulated across space and through time, beginning 4.5 billion years ago. The possibilities are endless.

This revolution will also change the way we interact with computers. We will interact with computers on our terms, not theirs. How will we treat these new species of computers that simulate human thought and attitudes? For example, will they have special rights?

The impact of the revolution in informatics and supercomputers on the biological sciences will alter our notions of life and existence. In genomics, it will allow genetic engineering to modify our bodies, our life spans, and what our descendants will look like. We may not like this, but this looks like the future. In neuroscience, it will give us the ability to understand detailed brain function, determine the nature of thought and consciousness, and even modify the mind and perhaps cure all mental illness. In bioscience, it will fulfill Harvard University's Ed Wilson's prediction that the coming century will be the century of the environment. Economic analyses and projections will become eco-economics. Health will be eco-health. Politics, natural resources, and energy will all have "eco" before them, and the quality of life will certainly be calculated as an "eco" property.

As a consequence, society too will undergo revolutions. First, there will be an even greater digital divide than there is now among nations. We will live in an expanding knowledge economy that is driven by information technology. Therefore, digitally poor nations will also be increasingly poor in knowledge and their economy.

As the realm of the unknown continues to shrink, especially in biology and cosmology, society will grapple anew with the meaning of life, the behavior of the mind, the purpose of human existence, and our sense of place and purpose in the universe, a realm currently occupied by religion and philosophy. We will arrive at a new sense of place and purpose in the universe.

## Biodiversity Science and Informatics

Biology is perhaps the most complex of sciences. It covers 20 powers of ten, from single nucleotides to genomes, from proteins and cells

to tissues and organ systems, from individuals and populations to species, communities, and ecosystems and, lastly, to organic molecules in space. Because of *Homo sapiens*, biology also encompasses human systems and their interaction with the biosphere. Rita Colwell, director of the National Science Foundation, calls this biocomplexity.

### *How Informatics Is Changing the Nature of Research in Biodiversity Science*

There are more than 1,000 natural history museums worldwide. They are biodiversity observatories, although we do not call them that. Collectively, they record 300 years of the biological exploration of this planet. They hold about three billion specimens of Earth's plants and animals collected during that exploration. They document more than two million species of Earth's plants and animals. And they archive the data on the global composition, identity, spatial distribution, ecology, systematics, and history of these two million species. They provide the raw research material for revealing the patterns, processes, and causes of evolutionary and ecological phenomena. These enormous biodiversity collections are an invaluable knowledge commodity. I use an economic term deliberately because, in the end, stewardship of global biodiversity and ecosystems is an economic necessity.

Natural history museums, then, are the libraries of life. What kind of data do they have, and how important is it to employ informatics to integrate this data with information? One example involves Mexican birds. The maps below show occurrence data for all Mexican birds. These data are housed in the British Museum in London, the Field Museum in Chicago, the Paris Museum in France, and the University of Kansas Natural History Museum in Lawrence, Kansas. If you combine all the data points on Mexican birds from 44 museum collections around the world, you have a "world museum" record of Mexican bird occurrences and a much denser, more robust database with which to conduct biodiversity science.

**Figure 1a**  
**Museum Biodiversity Data**  
**British Museum Mexican Birds**



**Figure 1b**  
**Museum Biodiversity Data**  
**Field Museum Mexican Birds**



**Figure 1c**  
**Museum Biodiversity Data**

**Paris Museum Mexican Birds**



**Figure 1d**  
**Museum Biodiversity Data**

**KU Museum Mexican Birds**



## Figure 1e Museum Biodiversity Data

“World Museum” Mexican Birds (44 collections)



In another example, the maps in Figure 2a–2d show data points for fishes in the Caribbean area from specimen records at the University of Florida, Tulane University, and the University of Michigan. Combine all three and you have a robust idea of fish biodiversity in the Caribbean and the Mississippi Delta.

How can we do this? How can we combine the information scattered among the world’s biocollections, these libraries of life, which, until now, have been “stealth” libraries?

### Species Analyst

At the University of Kansas Natural History Museum and Biodiversity Research Center, informatics research (funded by the National Science Foundation) has produced an open, free network called Species Analyst. This network allows any user to query simultaneously multiple collection databases worldwide via the Web. A user can retrieve the data as a dot map (or other analytical/visual tool) that depicts the known occurrences of that species based on authoritative museum collections.

**Figure 2a**  
**Museum Biodiversity Data**

**U Florida University Western Hemisphere Fishes**



**Figure 2b**  
**Museum Biodiversity Data**

**Tulane University Western Hemisphere Fishes**



**Figure 2c**  
**Museum Biodiversity Data**

**U Michigan Western Hemisphere Fishes**



**Figure 2d**  
**Museum Biodiversity Data**

**Three Museums Western Hemisphere Fishes**



How does Species Analyst work? It is a distributed query and retrieval system that begins with the specimen databases at museums worldwide that are hosted on Internet servers. These, in turn, respond to an information retrieval all-purpose interface linked to a desktop application. The user submits the query via the Web and receives an answer in seconds—a process that used to take months and years of poring through card catalogs.

However, being able to assemble and visualize where animals and plants live and have been collected according to museum records is not enough. We want to turn a descriptive science into a predictive science. We want to predict plant and animal ecological niches and distributions based on those data points. Therefore, we add an analytical component using environmental coverages—topography, climate, and so forth—and a predictive algorithm called GARP (Genetic Algorithm for Rule-set Production). (A computer scientist at the San Diego Supercomputer Center, which is a partner in this research, produced this component.) The analysis yields a predictive model of the geographic distribution of a species based on the data points from museum records. Essentially, Species Analyst builds an environmental-niche model for that species, which is the basis for calculating the species' probable geographic distribution in its native region.

The real power of Species Analyst becomes manifest in “what-if” scenarios. If the mean annual global temperature increases five degrees Celsius (which is predicted for the next 50 to 100 years), what would be the effects on biodiversity? We change the climate parameters in GARP to model a new, post climate-change distribution based on the environmental-niche model for that species. We can model the spread of an invasive species or emerging disease from one continent or region to another by overlaying the environmental-niche model of the species or its vector onto the non-native area that is being invaded.

Species Analyst also provides one-stop, point-and-click shopping for other properties of the particular species of interest. It enables the user to access Genbank for genomic data and Zoological Record for published information on that animal species. Species Analyst can also be launched with ArcView, which provides powerful geospatial analytic tools.

### *Applications*

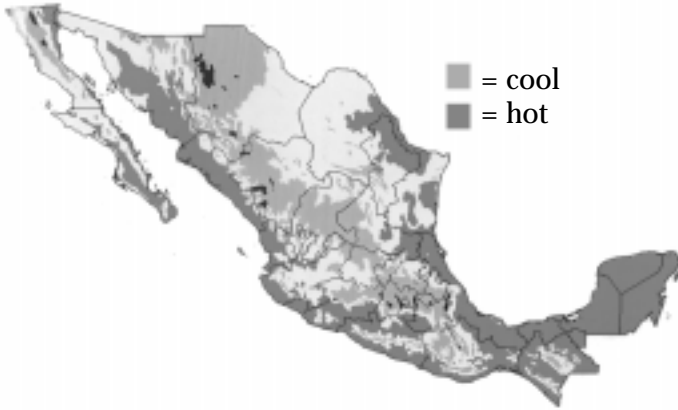
We can apply this knowledge in a number of areas.

**Biodiversity and climate change**

One area of application is simulating the effects of global climate change on animals and plant distributions, for example, on Mexican bird species. Figures 3a and 3b compare Mexico's current projected temper-

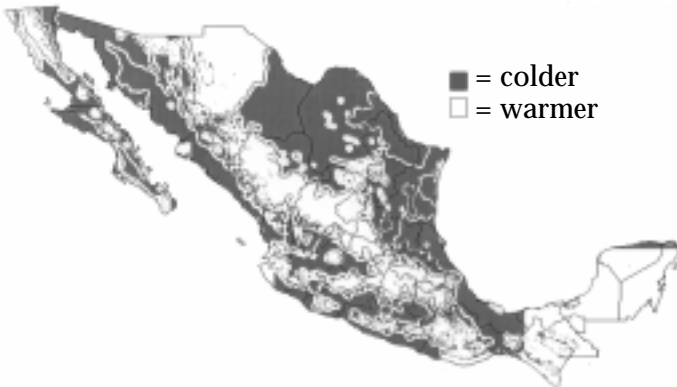
**Figure 3a**  
**Predict biodiversity: Climate change**

**Mexico's Current Temperature Regime**



**Figure 3b**  
**Predict biodiversity: Climate change**

**50 Years: Mexico's Projected Temperature Change**



ature regimes according to a conservative model of global climate change during the next 50 years. The darker shade indicates warmer areas and the lighter shade cooler. There are few cooler areas on the map, except for the mountainous area that runs through central Mexico.

### **Mexico's Green-striped Brush Finch**

This bird is well-represented in 43 museum collections. In seconds, Species Analyst retrieves and maps the data points from those collections and, using GARP, models the current geographic distribution of this species in Mexico (Figure 4a). When we plug in the climate change data for the next 50 years (from above), the predicted occurrence of the Green-striped Brush Finch (shown in dark areas in Figure 4b) is significantly reduced over its modeled current distribution (shown in Figure 4c). This is a powerful tool for conservation and natural resource planning for biodiversity and wildlands.

## **Figure 4a Predict Biodiversity: Climate Change**

### **Green-striped Brush-finch; Known occurrence in Mexico**



**Figure 4b**  
**Predict Biodiversity: Climate Change**

**Green-striped Brush-finch; Predicted Occurrence After Climate Change**



**Figure 4c**  
**Predict Biodiversity: Climate Change**

**Green-striped Brush-finch; Predicted Current Geographic Distribution**



### The West Mexican Chachalaca

Given current projections of global climate change, this bird's current predicted habitat (shown in Figure 5) will become highly fragmented within 50 years, resulting in a severely disjunct distribution that is the typical precursor of local, if not complete, extinction.

#### Informing conservation policy

Biodiversity informatics can play a powerful role in informing conservation policy. For example, 16 bird species are native to an area in southwest Mexico's dry forest. In one small area (shown in white in Figure 6a), 12 of these 16 species co-occur. In a second small area, farther west, four of the 16 species of birds co-occur. Where should the Mexican government invest in nature reserves to conserve native Mexican bird biodiversity? They declared the first area a reserve, but they then put a second reserve adjacent to the first, missing the second area of major bird concentration (Figure 6b).

### Figure 5 West Mexican Chachalaca

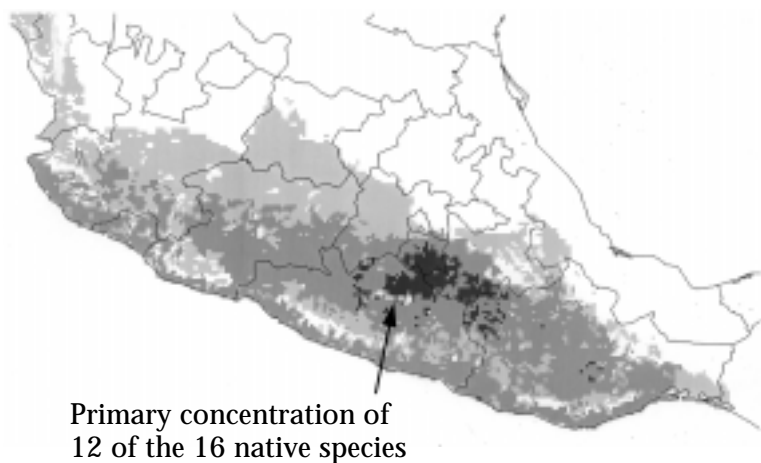
#### *Ortalis poliocephala*



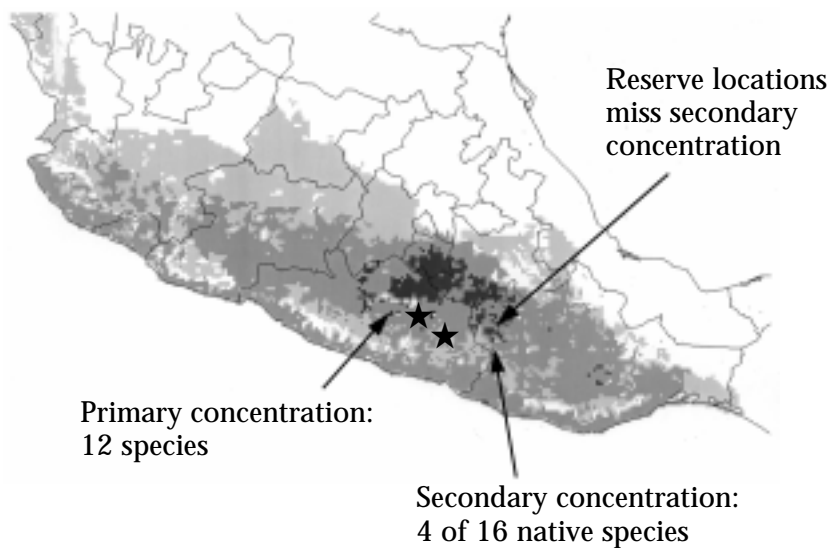
Before (light grey) vs. after (dark grey) climate change

- habitat fragmentation
- fragmented distributions

**Figure 6a**  
**Inform Policy: Conservation Areas**



**Figure 6b**  
**Inform Policy: Conservation Areas**  
**Where Should Mexico Invest In Conservation?**

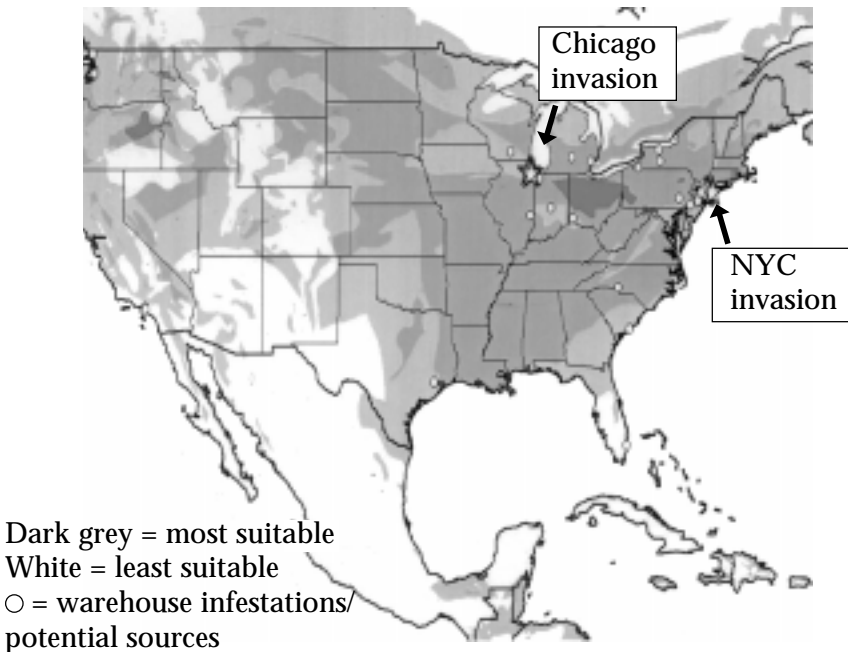


### Informing invasive species policy

Species Analyst can be a powerful tool in predicting the devastating spread of destructive invasive species. A good example is the Asian Long-horned Beetle, which invaded the United States a few years ago from the People's Republic of China in packing material. It is now eating its way through forests and parks New York and Chicago, destroying the stands of trees.

Based on just 40 museum specimen records of the Asian Long-horned Beetle, Species Analyst and GARP modeled the ecological niches and the likely distribution of this species in China. We added the influence of 18 climate variables in China, overlaid those on North America, and calculated a predictive model of the likely spread of the Asian Long-horned Beetle across the United States (dark area is the most suitable for invasion, white is the least in Figure 7). Using fire-spread algorithms, we can also model the beetle's dynamic spread across the United States.

**Figure 7**  
**Inform Policy: The Asian Long-horned Beetle**  
**Predict the Spread in North America**



### **A global biodiversity informatics network**

The Species Analyst network is global. It now has access to 50 million authoritative records of animals and plants and is growing weekly. The Global Biodiversity Information Facility, which was established in 1999 under the Megascience Forum of the Organization for Economic Cooperation and Development, has adopted Species Analyst as a model network for access, integration, and analysis of biodiversity information worldwide. Expansion to other museums and biocollections is constrained by the number of trained programmers we can hire.

What is the future? We are now putting the final polish on a screen-saver project that uses Species Analyst and will soon be available to the public. It has the working title of Lifemapper and, like the screen saver *seti@home*, uses the computer's available CPUs. It automatically queries and retrieves museum records of one of the two million known species of animals or plants on Earth, integrates the specimen point data, calculates its predicted geographic occurrence, and archives the model for future use in research and education. As new data for a particular species are added to museum databases, its predicted geographic distribution can be updated. Users can also designate a favorite group of animals or plants and Lifemapper will restrict its biodiversity modeling and prediction to those species. Users' names will be permanently associated with the predicted distribution of each species calculated on their machines, with a cumulative biodiversity "life list" to their credit. If 500,000 people register for Lifemapper, we can develop predictive environmental-niche models for all two million known species on Earth in one year.

### **Conclusion**

Returning to the grand challenges, what should we be doing about information across Earth systems, whether it is biodiversity data, ecosystem data, genomic data, geochemical data, geospatial data, geophysical data, climate and atmospheric data, or any other data related to the biosphere? First, the data need to be integrated in real time for simulation and predictive modeling of complex, Earth systems phenomena. Second, we need to provide this information in real-time global streams to all citizens as a public knowledge server for Earth systems information.

We need to do likewise for human systems—integrating political data, agroeconomic data, demographic data, population data, public health

data, socioeconomic data, land-use data, and all other relevant data for modeling and effect-prediction of human systems.

Finally, Earth and human systems information require integration for informed planetary management. Here is where the next generation of supercomputers, the “thinking machines,” will help design the future, because it is only this kind of brute artificial intelligence that will be able to turn such levels of complexity and chaos into reliable effect-predictions, prioritization of issues, and policy advice for our governments and citizenry.

### Acknowledgement

The following research scientists are responsible for developing the Species Analyst network, GARP, and Lifemapper: Drs. James Beach, Ricardo Pereira, A. Townsend Peterson, David Vieglais, Gregory Vorontsov, and Ed Wiley at the University of Kansas Natural History Museum and Biodiversity Research Center, and David Stockwell at the San Diego Supercomputer Center. Peterson and students conducted the research applications on Mexican birds. I am indebted to these individuals for discussions and for greatly enhancing my knowledge of biodiversity informatics.