BIODIVERSITY VIRTUAL RESOURCES INC.

Building a Distributed Knowledge Network

The Vision

In a small town in northern Ontario a group of senior high school students is collecting information about wetland habitats from their local pond. They use a digital camera to take pictures of the various plants and animals and a data-capture template to record their observations. Later that evening in the school's computer lab, they transfer the data and images to a computer connected to the Internet and upload them to the distributed Biodiversity Knowledge Network.

To add the images and observations to the knowledge base they use a simple tool that ensures data standards are adhered to and quality is assured. These students are not alone; there are hundreds of groups collecting data about wetland habitats and uploading observations to the network. Through the distributed Biodiversity Knowledge Network, scientists studying indicator species such as frogs, can search for and access datasets that include frog-related information. They find and download all the wetland habitat datasets and integrate them for analysis.

Analysis of the data collected by students across North America reveals changes to the frog population, indicating some climate change is occurring. Policies and programs are implemented to address the cause of the changes, reduce risks and raise awareness of how to limit the impact. Although the example above is fictional, the technology exists today to allow for the creation of such a distributed network, creating a virtual computing environment.

The Challenges

Researchers, naturalist groups, policy makers and governments agree there are fundamental challenges facing the biodiversity research community today. The root cause of these challenges is the continued use of 'centralized' data collection and management models.

From Centralized to Distributed

Costly to implement and maintain	→	Uses existing IT resources
Difficult to locate and access	→	Designed for Search and Retrieval of Data
Data collected using various formats and standards	→	Data Capture Tools ensure consistent use of data formats and standards
Data unavailable or lost when repository offline, shut-down or deleted	→	Data persists even if original data source is unavailable

Inability to Share Data: Using a centralized model, the data collected by the high-school students in our example would likely be stored in a makeshift database on a computer in their school. Scientists and researchers would never know about the project, let alone have access to the data.

Lack of Funding: Even if researchers knew of wetland habitat projects being conducted by schools across North America, acquiring and re-formatting each dataset to some standard would be costly and time-consuming. Furthermore, with the general lack of funding facing biodiversity researchers today, it would be difficult to justify and support the large investment in IT infrastructure required to store the data and run complex analysis.

Loss of Data: Because the students are unaware that their learning exercise could contribute to Scientists' better understanding of biodiversity and the state of the environment, they simply delete the data at the end of the term. The result is the loss of valuable information about wetland habitats in the region.

The Mission

At **BVR Inc.**, our mission is to provide flexible, easy-to-use tools for researchers and naturalists that will allow them to participate in the creation and development of a Distributed Biodiversity Knowledge Network.

These tools facilitate participation in the distributed knowledge network over the Internet by allowing participants to:

- Capture and manage data collections using standards-based methods and nomenclature
- Search for, access and integrate standardized data and information contained within the network
- Share information and collaborate with others on the network
- Protect data and information against misuse or loss

Using add-on modules *naturalists* can use the tool-set to create information based on their own data or that of others.

Researchers will use modules that allow for robust searching, in-depth analysis and computations, and multi-source data integration.

The true value of such a distributed knowledge network will be seen as the data contained within becomes complete. As this occurs, the biodiversity community will have one data source that contains a worldview of all the information available on the innumerable species that inhabit our planet.

In time, this network will be an invaluable source of biodiversity information that will ensure our continued health - physically, economically and spiritually.

BVR Inc. **+1.613.294.9014** **dpladouceur@bvrr.com** **http://www.bvrr.com**

The Distributed Knowledge Network Explained

With the advent of the Internet it is now possible to connect individual computers into a worldwide network. The next logical step is to provide a resource sharing mechanism.

Each computer in the network provides idle CPU cycles and a portion of its disk space. The result is a virtual computer comprised of potentially millions of nodes. Individual users may take advantage of the virtual resources for processing and storing data.

The architecture is comprised of:

- An Application Layer providing a rich set of user features;
- A Metadata Engine enforcing standard nomenclatures allowing for the creation of a virtual repository; and
- A Transport Layer allowing for communications between nodes.

Our technology provides users with the tools to create flexible data capture environments while enforcing strict metadata standards.

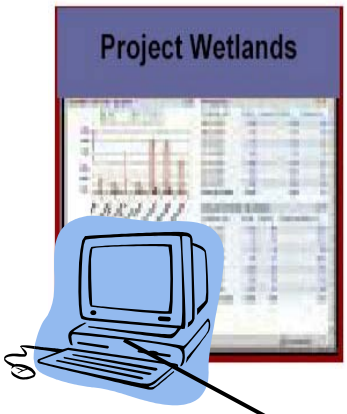
Data stored in the virtual repository are abstracted from the presentation. This allows for the data to be stored in a consistent, standard format, yet presented to each user in the format or language they wish.

Data contained in the virtual repository is physically stored on a multitude of individual computers. There is no need for central storage and backup since the data can be replicated within the distributed network, making it persistent.

The virtual repository takes advantage of the physical storage device of the individual computers in the network. A network of 1 million computers each contributing 1 gigabyte of disk space would create a virtual repository of 1 peta-byte or 1 quadrillion bytes of storage capacity. No central server architecture could provide such storage capacity.

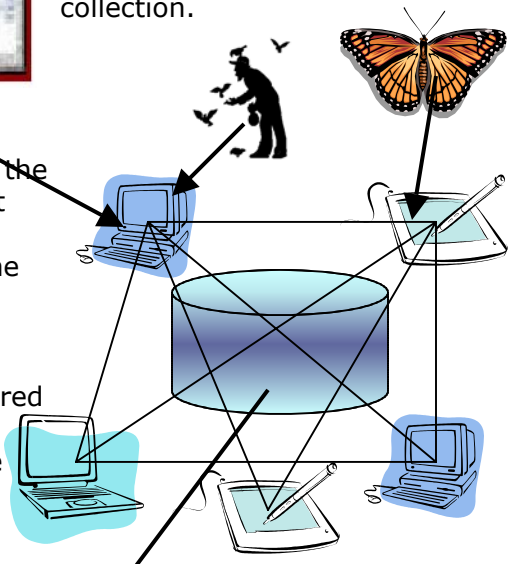
Quite simply, a distributed data repository addresses the challenges brought on by the use of centralized data repositories. Architected from the ground up to facilitate large-scale data collection, search, integration and analysis – distributed networks are the future.

Lets take a closer look at how a Distributed Knowledge Network works…



A scientist studying wetlands creates a project template and distributes it to individuals and organizations for data gathering in the field.

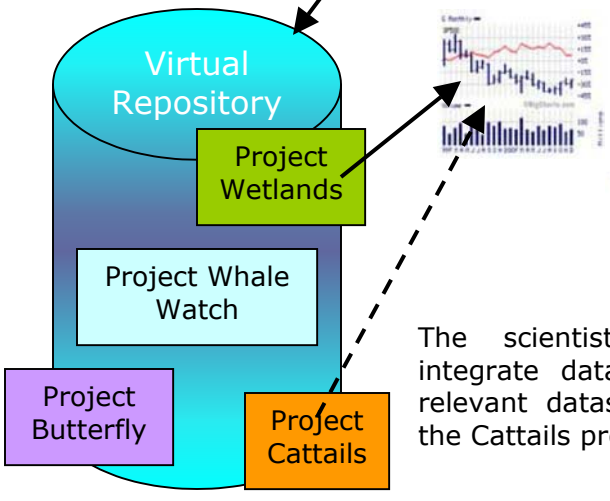
The template contains a data capture outline and the methodology for data collection.



Naturalists involved in the project can now collect data using the project template created by the scientist.

The observations from each individual are stored in a virtual repository accessible to all on the network.

At any time, the scientist can access the data contained in the virtual repository. Because the data has been collected using standard methods and stored in a common format, the multiple data sets are easily integrated, analyzed, and converted into information.



The scientist may also integrate data from other relevant datasets, such as the Cattails project.