Wildlife Health Monitoring Network

Data integration and exchange are vital when assessing the impact of wildlife disease globally …

Background

The National Biological Information Infrastructure (NBII) <www.nbii.gov> is an electronic information network that provides access to biological data and information on our nation's plants, animals, and ecosystems. Data and information maintained by federal, state, and local government agencies; non-government organizations; and private-sector organizations are linked through the NBII gateway and made accessible to a variety of audiences including researchers, natural resource managers, decision-makers, educators, students, and other private citizens.

Implementation of the NBII is being accomplished through the development of nodes that serve as entry points to the NBII and the information held by partners. These nodes function as fully digital, distributed, and interactive systems that focus on developing, acquiring, and managing content on a defined subject area or a geographic region. The NBII Wildlife Disease Information Node (WDIN) addresses the need for information on a variety of wildlife diseases and their implications, including those affecting wildlife, domestic animals, and humans. WDIN operations are managed through the collaboration of its major partners: the USGS National Wildlife Health Center and the Gaylord Nelson Institute for Environmental Studies at the University of Wisconsin-Madison.

## Goal of the Wildlife Health Monitoring Network

Integrating different types of wildlife disease surveillance data together provides critical information for understanding wildlife disease patterns and their potential impact on wildlife, human, and domestic animal health. To help address this need, WDIN is building the Wildlife Health Monitoring Network (WHMN), a Web-based open source system with interchangeable modules that support data entry, storage, reporting, analysis, and exchange in collaboration with many partners. The goal of this system is to create tools for institutions and individuals to track and store wildlife health data in a way that is useful and meaningful to them, as well as allowing them to share their data, as they deem appropriate, with the rest of the Network.

Flexible Database Structure

At the core of the WHNM development, WDIN's collaboration with the University of Wisconsin's Division of Information Technology has produced a flexible, scalable, and robust database schema. The database design allows flexibility for evolving tracking requirements by representing semantic structure through data rather than though table structure. WDIN's approach to database structure has been to replace traditional table fields describing a measurement or similar record (e.g., location, age, sample), with a concept for each data entity having a type and a value. To record new types of data, the database only requires an added new type to the table.

Contrast this with a more traditional approach of adding a new column to a table and then retrofitting data input and reporting mechanisms to accommodate the newly hardwired data element. The approach of soft coding types allows database users to record new types of data without needing information technology support.
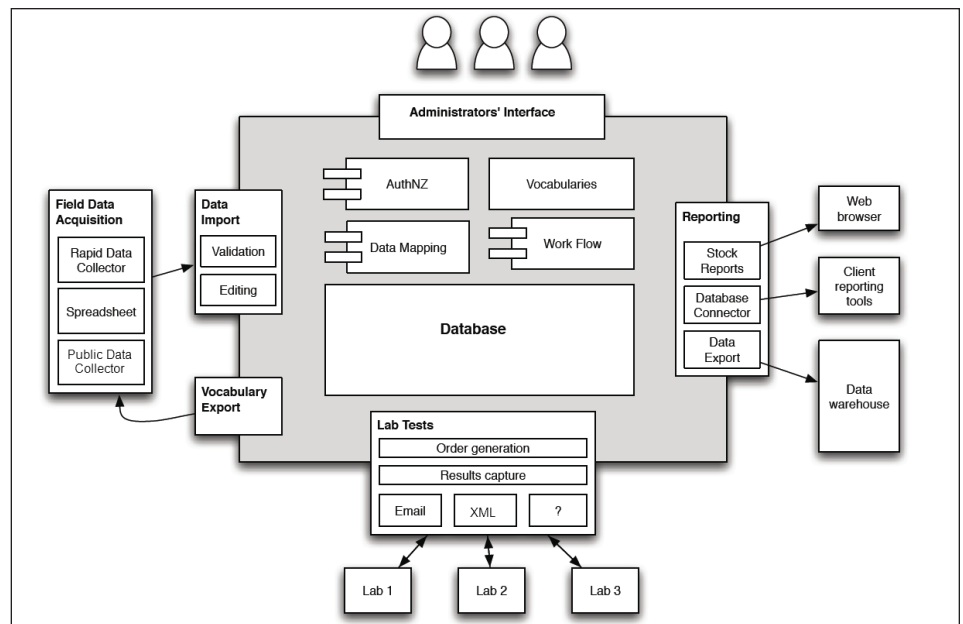
The loosely-coupled architecture depicted allows WDIN to include or swap functional components easily without re-engineering the application's core.

Adaptable Data Entry

To provide a means to standardize data entry, while being flexible enough to allow for individual organization's needs in regards to data collection and entry, WDIN anticipates offering tools which would allow user groups to design their own Web-based forms. These forms could be unique in look, feel, and contents, but have the ability to draw from standardized questions and answers to control consistency from application to application and, therefore, help with data comparison in the future.

Additional data import functions would also allow for Microsoft Excel files or other commonly used data storage formats to be accommodated within WHMN. WHMN's ability to handle other data formats lies in the heart of the next topic, data integration.

Innovative Data Integration and Exchange

WHMN includes a module for moving the data from the entry form to the database. This same tool will enable data from other systems to be "mapped" to the WDIN local database, allowing for significant ease of data warehousing to improve creating reports and enhancing the ability to do analyses. Each data source would have a customized ETL (Extract, Transform, Load) function generated that would crosswalk the data elements from the source database to the appropriate observations in WHMN and also map the terminology used from the native terms to the centralized terms. This process will allow a variety of source data sets to be viewed and mapped together.

Data need not originate within the WHMN (e.g., use the Public Data Collector) to be shared with the WHMN database. The ETL process can be set up to query against external databases, where the originating data and information reside. This ability allows external partners who wish to make data available to the WHMN system to continue to use their existing data structure, and can be accomplished without actually transferring the data.

Creating Maps

Visualizing information spatially on maps provides a key component
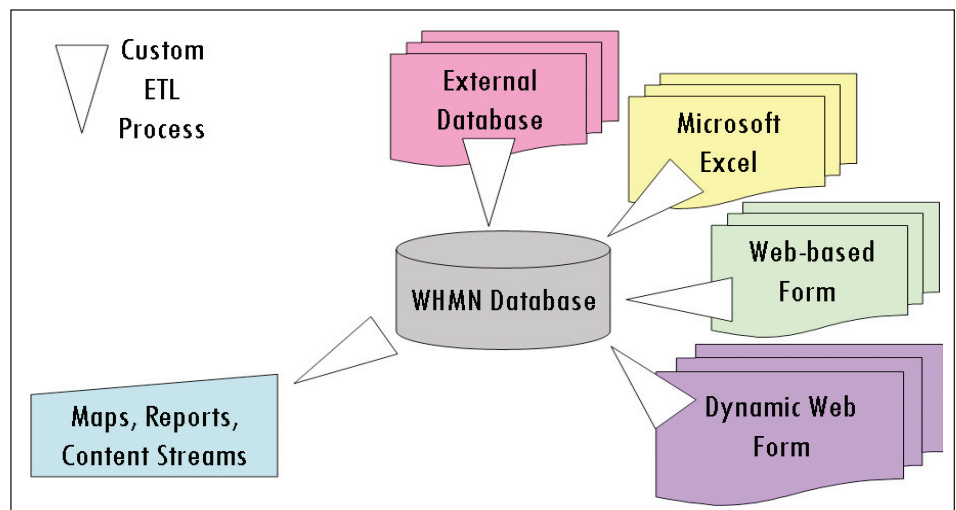
The source-specific, XML-based Extract, Transform, and Load (ETL) processes allow unique data sources to be mapped onto the WHMN database structure; users will then be able to generate reports, maps, or exports from individual data sets or the entire collection of information.

to explore data. Using interactive mapping software integrated from the Pentaho Business Intelligence (BI) package, WHMN will allow users to view the data geographically. Functions within the package allow users to do basic filtering of all visible data sets through an easy-to-use Web-based interface.

Building Formulated and Specialized Reports

In addition to providing the mapping functionality of the WHMN database, the Pentaho BI also allows WHMN to serve up pre-established reports, which are dynamically refreshed each time they are requested. Users can also generate their own reports using the ad-hoc query tool built into the Pentaho package. Administrators can set up data views and allow users to put together their own queries without any help from technical staff.

Open Source Applications – Tools for All

The goal of the WHMN toolset strives to use as many open source technologies as possible to ensure that the application can be set up in as many instances as desired and keep a low cost overhead. In addition to adopting open source tools in existence for report generation and mapping (Pentaho BI), the database backend utilizes PostgreSQL, an open source enterprise class database with 15 years of development history. By adopting these widely used open source tools in a loosely coupled

environment, WHMN can insert or substitute tools as new system needs are identified in the future

Looking to the Future

As data sets are identified and integrated with the WHMN, additional analysis components will be investigated and incorporated into the system beyond simple reporting, querying, and mapping. Additional tasks for future expansion include workflow management functions, laboratory integration mechanisms, and additional terminology development with stakeholders and partners.

Using this innovative, flexible system, the Wildlife Health Monitoring Network, researchers, and decision-makers – wildlife and non-wildlife related – who share concerns about wildlife disease can coordinate disease control and prevention efforts within and across political and disciplinary boundaries in effort to protect the well-being of all living things and promote a healthy ecosystem.

For More Information

Jennifer Carlino
Node Manager
USGS NBII Program
Phone: 303-202-4260
E-mail: jcarlino@usgs.gov

F. Joshua Dein
Principal Investigator
USGS National Wildlife Health Center
Phone: 608-270-2450
E-mail: fjdein@usgs.gov