ASSEMBLING THE KNOWLEDGE BASE

Dorthe B. Carr, J. Mark Harris, Lisa K. Wilkening, Christopher J. Young, Matthew N. Chown, B. John Merchant, and James R. Hipp

Sandia National Laboratories

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. DE-AC04-94AL8500

ABSTRACT

The National Nuclear Security Administration (NNSA) Knowledge Base (KB) is a combination of the information content, database storage framework, and interface applications needed to provide integrated research products in a form that will allow the United States National Data Center (USNDC) to meet its monitoring objectives. To build a successful Knowledge Base, all the different pieces must be validated, integrated, tested, and organized to make a seamless, high-quality product, which is effective and easy to use. Sandia National Laboratories (SNL), as the KB Integrator, is responsible for accomplishing this.

Component integrated research products were delivered to the KB Integrator by KB Contributors for the latest release of the KB. These products were made up of a combination of one or more related datasets. Requirements for the datasets and the accompanying documentation were established early in the process by the KB Working Groups and given to all contributors. At the KB Integrator, the integrated research products were organized into five categories, based on their format, and then processed. The five categories are event data, parametric grid data, contextual data, research tools and supplemental information. For event data, the format is Oracle database tables (which follow the established NNSA Knowledge Base schema) and complete integration was achieved using the Lawrence Livermore National Laboratory (LLNL) program Orloader, and various SQL scripts. For parametric grid information, the format is a custom database format known as a PG database, and no further processing was needed as integration of the various PG databases was not a goal. For contextual data, the formats are shapefile (vector) or imagery (raster), and some processing was required to tailor the geographic extents of the data and to make it all accessible from a single custom ArcView GIS (Geographical Information System) Project. For research tools – i.e. software used to help create and access other integrated research products – there is no standard format; a researcher can use the language of their choice, as long as they provide instructions for compiling the program and information about the environment under which the tool was developed. Most of the integration work was accomplished during the development process for the tools because tool designers worked with the KB Integrator throughout the process to make sure the required hardware and software needed to run the tool would be available at the customer's site. Finally, for the supplemental information (white papers) the established format was PDF (portable document format) documents, and the only processing needed was to convert the original format to PDF if necessary.

Once the products were all processed, they were then organized and indexed for ease of access and the validation process began, using a new process that was developed for this release cycle. The integrated research products were divided among a dozen "Validators" at SNL. Each Validator was tasked to make sure all the components of the integrated research product had been delivered and to run a functional test on the product provided by the researcher. The functional tests were done twice. The first time was as a stand-alone test on the product before it had been integrated with other products in the Knowledge Base. The tests were run a second time after the products had been integrated to insure that nothing had been altered during the integration process. Once integration and validation had been accomplished, an interface to the Knowledge Base was created using the KB Navigator, a new piece of software developed by the KB Integrator specifically for this purpose. The KB Navigator allows users to find products, check geographical extents, access standard GIS datasets, view metadata and launch the research tools. The Navigator is supported by a custom database schema, which contains details about every product in the NNSA KB. As a final test that the NNSA Knowledge Base had been successfully built, the KB Navigator was then used to browse and examine all the delivered integrated research products.

OBJECTIVE

The National Nuclear Security Administration (NNSA) Knowledge Base (KB) is a combination of the information content, database storage framework, and interface applications needed to provide integrated research products in a form that will allow the United States National Data Center (USNDC) to meet its monitoring objectives. To build a successful Knowledge Base, all the different pieces must be validated, integrated, tested, and organized to make a seamless, high-quality product, which is effective and easy to use. Sandia National Laboratories (SNL), as the KB Integrator, is responsible for accomplishing this. In this paper, we describe the process used by the KB Integrator to put together the latest release of the NNSA Knowledge Base.

RESEARCH ACCOMPLISHED

Preparing for Integration

The KB Integrator, as defined in Gallegos et al. (2002), is a group of individuals in the Ground-based Nuclear Event Monitoring Research and Engineering (GNEM R&E) program at SNL, who combine Information Products (IPs) into the NNSA Knowledge Base. The process involves verifying and functionally validating all the documentation and products that are submitted, integrating them together and organizing the information so that it is easy to access and use. During prior releases, two or three people at SNL were designated to do the verification and functional validation of the submitted integrated research products. This did not work well, because the total number of submitted integrated research products ranged between 70 and 80, and it was very difficult for such a small group of people to do both verification and functional validation in the time allotted for that process.

For the latest release of the KB, researchers, known as Scientific Integrators, were given approximately ten months to create their integrated research products and write draft documentation, before it was due to the KB Integrator. The KB Integrator then had two months to verify, functionally validate and integrate the products into the KB. Because of the problems that had occurred in integrating prior KB releases, SNL redefined the internal process of how integration of the KB was done. The first was to name one person within the KB Integrator to coordinate the integration and be the point-of-contact for the Scientific Integrators.

Four months before the integrated research products were due to the KB Integrator, we defined a directory structure and database accounts to hold the integrated research products. The directory structure that was developed consisted of three major directories: Documents, Tools and Products. The Documents directory was designed to hold the IP documents. The Tools directory was designed to hold the research tool products. It provides a framework that facilitates code maintenance while enabling extensibility and allows different tools to share codes and libraries. The Products directory grouped integrated research products by IP. This is a convenient way to keep track of the individual integrated research products during integration. For the research tool products, we planned to have the products reside in the Tools directory, with appropriate links back to the Products directory. Once the directory structure was finalized on the unclassified server at SNL, it was replicated in the SNL vault and at the customer site, the Air Force Technical Applications Center (AFTAC).

We also decided that we needed to get more people involved in the verification and functional validation of individual integrated research products. Each expected integrated research product was assigned to a Validator at SNL, who was tasked with verifying and functionally validating that product. There were a total of 67 expected products and fourteen SNL Validators. Most Validators were assigned three or four products, and we tried not to match products to "experts" at SNL. Our goal was to try and catch problems that an inexperienced user might find. The first thing the Validator did was to make sure that all the entities (files, database tables, etc.) that comprise a product made it to SNL. The second was to run a test provided by the Scientific Integrator to make sure that the integrated research product could be accessed and used as the Scientific Integrator intended. This was not a test to look at the scientific validation of the product; instead it was considered to be a functional validation test. The Validator was to functionally validate the product first as a stand-alone product, and then run the same test again after all the products had been integrated. The KB Integrator coordinator would track the progress of all Validators and provide support wherever needed.

Six weeks before the integrated research products were due at SNL, the Scientific Integrators were required to submit the "Files" and "Installation Testing Procedure" sections from the integrated research product documentation. We requested all Scientific Integrators to make these two sections as complete as possible. The "Files" section contained a complete description of all the entities that made up the product. The "Installation Testing Procedures" section described the test the Validator was to run to make sure the product could be accessed and used correctly. Scientific Integrators were encouraged to send test data and their results so that the Validator could compare the results of the test done at SNL to the results of the test done by the Scientific Integrator. The Validator reviewed the two sections for the products to which they had been assigned. If they found any problems or did not understand the testing procedures, they contacted the Scientific Integrator and worked with them over the six-week period before the products were due to make sure problems were resolved.

Integration

Products arrived at SNL on January 21, 2003 and the integration process began (see Figure 1). The KB Integrator coordinator received the products and put the data into the appropriate place in the Products directory. If the product was unclassified it was put onto the unclassified server at SNL; if it was classified, it was put directly into the vault. Once a product was in place, the Validator was notified that his product was ready. If there were files or database tables the Validator could not find, he first went to the KB Integrator coordinator to make sure that all the information delivered for a product had been put in the correct place, or to make sure they were looking at the correct database account. If it was determined that not all the files listed in the "Files" section arrived at SNL, the Validator contacted the Scientific Integrator to get the additional files delivered to SNL. In some cases, there were additional files delivered in an integrated research product that had not been described in the "Files" section of the documentation. In these cases, the Validator contacted the Scientific Integrator to determine if these files were part of the product.

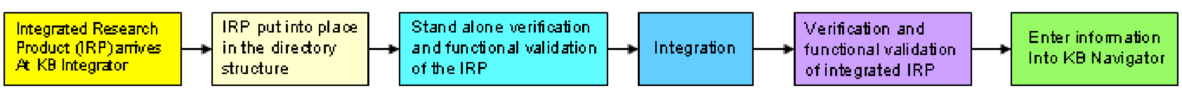

Figure 1. The process of assembling the Knowledge Base

Once the product had been verified, the Validator ran the installation test on the product. The first step was to do unit testing on the product, i.e. testing it as a stand-alone product. If the Validator had problems running the unit test, they contacted the Scientific Integrator for the product to try and resolve the problems. Most of the problems the Validators ran into with the installation tests were because the test was not well-defined, well-documented or the test sets that were to be used to verify the results of the test were not supplied to the KB Integrator. A README file was started by the Validator for each integrated research product that described how the test was run and any changes the Validator had to make from the Installation Test Procedures provided by the Scientific Integrator to get the test to run correctly.

As soon as an integrated research product had been both verified and functionally validated by the Validator, he informed the KB Integrator coordinator that the product was ready for integration and the KB Integrator coordinator moved the product into the SNL Vault (classified products were installed directly in the vault). The integration of an integrated research product depends on the format of the product. There are five different formats, depending on the kind of information – event data, which is stored in Oracle database tables, parametric grid information, which is stored in a custom database format, contextual data which is stored as shapefiles or imagery, research tools, which had no standard format, and white papers stored in PDF format. The bulk of the integration done by the KB Integrator was for event data and contextual data. Details on the integration process for each of these are given below.

Event Data Integrated Research Products

All event data is stored in Oracle database tables, with associated instrument response files and waveform files. We use the NNSA core table schema (Carr, 2002) for the basic information and the NNSA custom table schema for storing additional event parameters. The NNSA core tables separate into two groups: Lookup and Primary. The Lookup tables change infrequently and are used for auxiliary information used by the processing. The Primary tables are dynamic and consist of information that can be used in automatic and interactive processing. The integration of the event data fell into four categories: Lookup tables, Primary tables-catalogs, Primary tables-waveforms, and custom tables. We use the convention in the following paragraphs that table names are in **bold** and column names from the tables are in *italics*.

The Lookup tables consist of six tables that describe station parameters (**site**, **sitechan**, **instrument**, **sensor**, **network** and **affiliation**) plus the instrument response files (Figure 2). Most of the information on the station parameters came from Lawrence Livermore National Laboratory (LLNL), but both Los Alamos National Laboratory (LANL) and the AFTAC also provided station parameter information. At SNL there were some diagnostics run on the station parameter information to check for consistency problems, such as stations without responses, bad calibration periods (calper), etc. Problems found were resolved through communication with the information sources. The LLNL information was loaded in as the basic information, and changes to that information were only made when it was deemed that the LANL or AFTAC information was better. Then the station parameter information was changed from the NNSA convention to the convention used at AFTAC. The major changes involved making sure the networks, channel names and stations names matched the names used at AFTAC. At SNL, all the instrument response files were converted to frequency-amplitude-phase (FAP) format, which was the format requested by AFTAC. For each response file that was converted, we made sure that instrumentation calibration information (calibration factor and calibration period) were correct, and matched the calibration information in the waveform table (**wfdisc**).

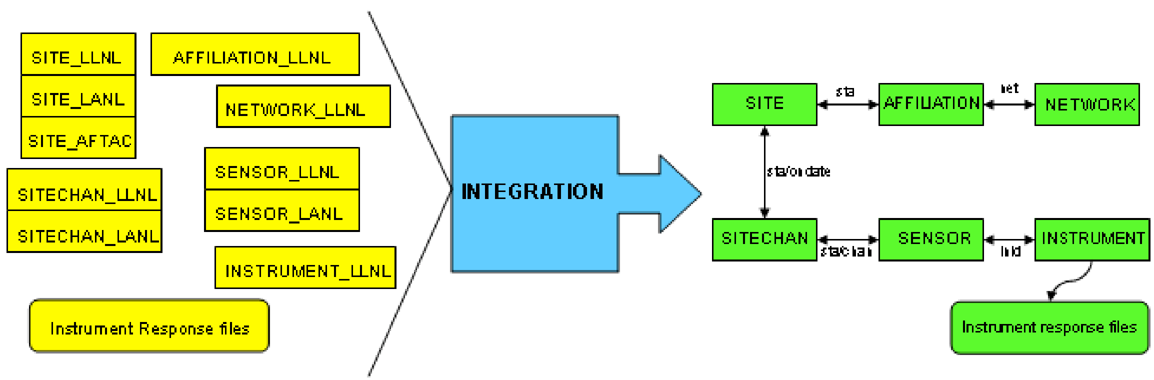

Figure 2. Integration of the Lookup tables

There are seven Primary tables that contain earthquake catalog type information: **origin**, **event**, **origerr**, **arrival**, **assoc**, **netmag** and **stamag**. LLNL and LANL provided global, regional and local earthquake catalogs to SNL as flat files or database tables. For AFTAC's use, it is essential that these catalogs be merged to a single set of tables and corresponding events in each be tied together. The catalogs were merged together by the KB Integrator using a program developed at LLNL called Orloader. As each catalog is merged (Figure 3), event locations common to multiple catalogs are assigned a common event identifier (*evid*) and unique origin identifiers (*orid*) in the **origin** table. The **event** table lists the preferred origin for each event based on a user-specified author preference table (**origin_authors_rank**) that ranks the catalogs so that catalogs that provide the best location are ranked higher. Any information from the catalogs that fit into the **origerr**, **arrival**, **assoc**, **netmag** and **stamag** tables are put into the appropriate tables. A **remap** table is created in the process of merging the catalogs that lists the *orids* and *evids* in the origin table and maps them to the original values (*origorid*, *origevid*). SNL started the merging process in December, 2002 with the global catalogs. Once the merging of global information was completed, the regional and local catalogs were merged into the tables. The entire process of merging all the earthquake catalogs took approximately two months to complete. All available catalog information as of January 21, 2003 was merged into the KB tables.

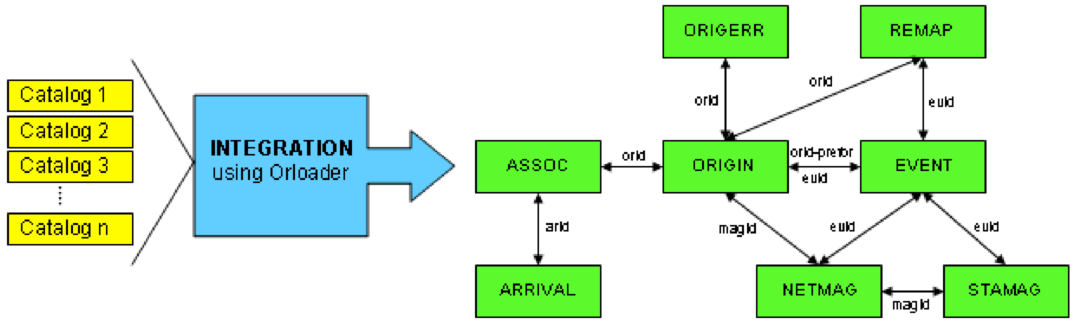

Figure 3. Integration of the Primary tables that contain earthquake information

The **wfdisc** and **wftag** Primary core tables contain information about the waveforms. The waveforms themselves must be in one of several supported binary formats. Both LANL and LLNL delivered **wfdisc** and **wftag** tables with their associated waveforms. To merge the information from the two laboratories, we first compared the two **wfdisc** tables to see if there were any identical waveforms (see Figure 4). There turned out to be 8 identical segments (out of ~282,000) and the labs were contacted to determine which information should be used for those segments. Then the waveform identifiers (*wfids*) in each table were remapped to the KB numbering scheme, using the **remap_wfids** table created by the KB Integrator. Once the *wfids* were changed the two **wfdisc** tables from the laboratories, were combined to make the KB **wfdisc** table. The same remapping of wfids was done in the two **wftag** tables from the laboratories. In addition, the column *tagid* in the wftag tables needed to be remapped. The *tagid* in the wftag tables for the KB are always *evids*. Using the remap table created in the process of merging catalogs, the *tagids* were remapped to the *evids* for the KB. Once the *wfids* and *tagids* were changed in both tables, the two wftag tables from the laboratories were combined to make the KB wftag table.

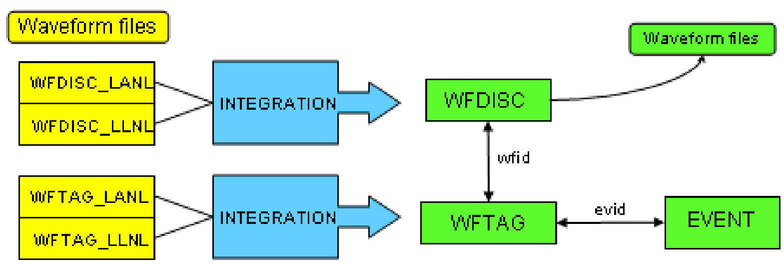

Figure 4. Merging the wfdisc and wftag tables from the two laboratories

For this release of the KB, there were 32 custom database tables. Five of these tables had one or both of the columns *orid* and *evid* that needed to be remapped to match the ids in the merged catalog tables. There were three tables we used to remap the *orids* and *evids*: the **remap** table created when merging the catalogs with Orloader, and the **remap_lanl_events** and **remap_llnl_events** tables which we had to create by hand using the merged **origin** table and origin tables sent from LANL and LLNL, **origin_lanl** and **origin_llnl**. The **remap** table created when merging the catalogs with Orloader could be used to remap *orids* and *evids* for the origins that came from the regional catalogs that were sent to SNL as database tables, because the *origorid* and *origevid* values in the **remap** table could directly be related to *orids* and *evids* in the origin information delivered from the laboratories. However, for the origins from the global catalogs and the regional catalogs that were merged from flat file information, the *origorid* and *origevid* values in the **remap** table related to the flat file numbering and not the numbering in either the **origin_lanl** or **origin_llnl** tables. To remap the origins from those catalogs, we created the tables

remap_lanl_events and **remap_llnl_events** that mapped the *orids* and *evids* from the merged **origin** table to the *orids* and *evids* in either the **origin_lanl** or **origin_llnl** table where the origin information (*lat*, *lon*, *depth*, *time*, *auth*) matched. Once the **remap**, **remap_lanl_events** and **remap_llnl_events** tables were in place, the *orids* and *evids* were changed in the **discrim_data**, **gttable**, **magnitude**, **nnsa_amplitude** and **nnsa_amp_descript** tables using simple update SQL commands.

In addition to remapping *orids* and *evids*, there were two tables where the column *arid* needed to be remapped and one table where the column *wfid* needed to be remapped. To remap the *wfids* in the **nnsa_amp_descript** table, we used the **remap_wfids** table that had been created for remapping *wfids* in the **wfdisc** and **wftag** tables, and changed the *wfids* using a simple SQL command. Remapping the *arids* was a little more complicated. All the arids that needed to be remapped came from LANL. Using the **arrival_lanl** table that LANL provided, we created the **remap_arids** table that mapped the *arid* from the merged **arrival** table to the *arid* in the **arrival_lanl** table by matching the arrival information (*sta*, *time*, *iphase*, *auth*). Again, a simple update SQL statement changed the *arids* in the **nnsa_amplitude** and **nnsa_amp_descript** tables to the KB numbering scheme.

The final remapping that needed to be done concerned identifiers only found in the custom tables: *fdid*, *corrid, paramsetid*, and *windowid*. We needed to make sure that before we combined tables from both laboratories that these identifiers did not overlap. If they did overlap, we added a constant to one laboratories numbers and then combined the information into single tables. The tables combined in this manner were **mdac_fd**, **mdac_fi**, **discrim_param** and **discrim_data**.

Once the integration was completed, all the database tables were put into a single database account and primary keys and indexes were set on the tables in consultation with AFTAC. The README files for the integrated research products that contained event data were updated to reflect the name of the database account.

Parametric Grid Data Integrated Research Products

The parametric grid (PG) data format is an NNSA custom format developed to store a wide variety of types of geophysical information that is not handled well by any single existing system: travel time tables, kriged correction surfaces, tessellated correction surfaces, layered Earth models, etc. PG data is stored as a set of binary files, which we refer to as a PG database. The PG software provides the functionality to store and retrieve PG objects from the database in a generalized manner, and is used by many of the software tools delivered with the KB. Each object stored in the database is decomposed into a Meta-Data Key (MDK) and a Serialized Binary Stream (SBS). The MDK is used to contain summary information about the object including its object type (class name), date of creation, location in the database of the associated SBS, pertinent attributes (station, phase, base model, frequency, etc.), a description string, and a unique 16 byte (128 bit) hash string, called the Key String Identifier (KSI). The KSI uniquely identifies an object's MDK for which it was created. Additionally, the MDK contains all dependency KSIs that are required to build the object associated with the MDK. All MDK objects are loaded from the database immediately after connecting to a database platform. More information about PG can be found in Hipp et al. 2002.

NNSA developed the Data Manager (DM) package to browse and merge PG databases created by Scientific Integrators, but for this KB release no merging of PG databases was done. From a software perspective, the merging is straightforward so we were able to code it into DM, but from a scientific perspective, it raises many troubling questions. For example, if travel correction surfaces are developed for the same station and phase by each lab, and their geographic extents overlap, what should happen in the region of overlap? The NNSA researchers collectively decided that such issues should be worked out before the PG databases were created, with the labs working together to create a single correction surface whenever there was an overlap. Thus, when the products were delivered to SNL, there were no regions of overlap, so merging was not needed. We kept the component PG databases from each lab and delivered them as such.   The various PG databases were placed in the directories for the corresponding integrated research products.

Contextual Data Integrated Research Products

Contextual data are generally stored in one of two models, vector and raster. For the purposes of the KB, vector data are stored in ESRI shapefile format for use with the ArcView Global Project, the KB's geographic information system (GIS). The vector data are made up of a series of features with explicit geometry. Examples of vector data in the KB include political boundaries (polygons), mine locations (points), road and railroad networks (lines), geologic provinces (polygons), and seismic station networks (points). Raster data are represented as cells of a uniform size, each containing a single numeric value. Most raster data are stored in some imagery format, such as TIFF, JPEG, etc. The KB also supports the use of ESRI Grid format raster data, which are generally used to represent surfaces, such as topography, bathymetry, or correction surfaces, models, etc., derived from integrated research products submitted to the KB. Although each dataset is unique, the processing of these data is similar. The general steps taken to process the data are described here and depicted in Figure 5.

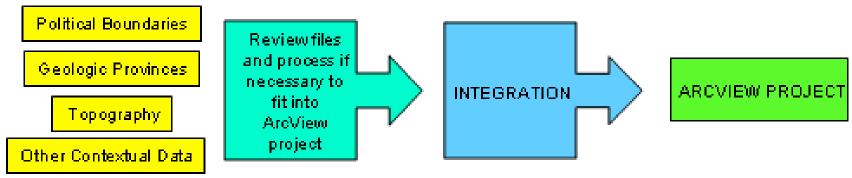

Figure 5. Integration process for Contextual Data

In most cases, the contextual data has been preprocessed before it reaches the KB Integrator. For example, the original source (shapefile, report with latitude and longitude coordinates, etc.) for a mining dataset may have covered more geographic area than is of interest to the submitting laboratory or customer. The Scientific Integrator will have clipped out the appropriate portions and submitted this subset to the KB Integrator. The Scientific Integrator will also have submitted an accompanying ArcView legend file, which is used by the ArcView software as a guide for symbolizing the dataset when it is displayed in the ArcView project. The same may be true for raster data, with the exception of imagery, for which ArcView does not support the use of customized legend files.

Once the KB Integrator receives the contextual data, a two-step process is performed. First, the Integrator reviews the data itself, checking that the files are not corrupted or missing, that they can be viewed in the ArcView application, that a legend file has been provided, and that the dataset and its attributes are consistent with the documentation describing the data. In some cases, the data must be processed further in order to conform to the ArcView project. This may involve the removal of attributes or inconsistent features from the dataset, “clipping” the data to correspond to specified geographic regions, or modifications to the legend file.

Second, the KB Integrator must incorporate the contextual data into the ArcView Global Project. This project is a customized application built specifically for the customer. It incorporates all contextual data submitted to the KB, as well as customized tools, functionality, maps, specialized database query tools, and custom interprocess communication procedures for communicating with other KB tools. The new datasets must be folded into this project in such a manner that it is both easily accessed by the user and complements the function of the project as a whole.

For inclusion in the ArcView Global Project, the data must be incorporated into a custom menuing system. The data are organized into specific locations on disk based on their function and geographic location, which is mirrored in the ArcView project. Rather than rely on a continuous connection to a relational database, data are organized into a lookup table that is stored within the project. The lookup table stores information about the dataset, including its location on disk, the location of its metadata, symbology information, and keys used by ArcView to facilitate access to the data itself. The table is utilized by the menuing system to access the data on disk, load the data into a View (or Map) in the project, and pull in any associated information stored with the data in the table record. In addition, the data may also be accessed by certain custom tools in the ArcView project. In all cases, however, the data, its symbology, and its metadata must be reviewed in the context of using the ArcView Global Project in order for the GIS as a whole to be tested. Once this process is concluded, the contextual data and the ArcView Global Project are considered complete and ready for submission.

Research Tool Integrated Research Products

There is no standard format for research tool integrated research products. The tool designers could pick any format they wanted, but the KB Integrator required that during development the designers needed to work with SNL to make sure that the required hardware and software to run the tool would be available at AFTAC for integration and testing. In past deliveries, tools were treated as autonomous packages despite the fact that many of them share libraries or even call one another directly, and this led to problems with the delivery and testing (duplication of files and general confusion). For this release, all SNL tools were installed as sub-directories into a directory called GNEM_Tool_Root. Links to executables for each tool were placed in a common bin directory also under GNEM_Tool_Root, and links to all of the libraries were placed in a common lib directory. This structure was very successful; it made it easier for the SNL tool developers to develop their products, it made the overall tool delivery more space efficient, and it also made it much easier to configure and test the tools after delivery. We integrated fourteen research tool products.

White Paper Integrated Research Products

The established format for the White Paper integrated research products was PDF. The Scientific Integrator sent to SNL papers either in PDF, or in the original format. For those that came in the original format, we changed the cover page to match the standard created for the KB, and then we converted the document to PDF. If the paper came in PDF, then we sent a copy of the standard cover page to the Scientific Integrator and asked him to include it in the document and resubmit the document in PDF with the new cover page attached. The various White Papers were placed in the directories for the corresponding integrated research products. There were six white papers in the KB.

Post Integration and Organization

As soon as an integrated research product was integrated into the entire KB, the KB Integrator coordinator informed the Validator so that he could verify and functionally validate the integrated product. By this point the verification was a fairly simple process; the goal was to make sure that all the entities associated with the product had been successfully integrated into the KB. The functional validation was again done by running the installation test that the Scientific Integrator provided. The goal was to make sure that the integration process did not corrupt the product so that the installation test could not be run successfully. If there were problems with the installation test, the Validator first worked with the KB Integrator coordinator to make sure they did not arise because of the integration and hopefully resolved them. If they could not resolve the problems, the Validator then went back to the Scientific Integrator for help to determine the solution.

For the latest release of the KB, the KB Integrator designed a graphical interface to easily organize and access the information in the KB. This tool is called KB Navigator. It is a "window" into the KB that allows a user to find the integrated research products, check geographic extents, access standard GIS datasets, view metadata and launch the research tools. An Oracle database schema was designed for the KB Navigator to hold all the pertinent information about the products. As Validators were verifying and functionally validating the integrated research products, the KB Navigator developers were entering information into the schema.

Nodes in the Navigator were created for each integrated research product and their associated datasets following the directory structure that had been set up prior to the start of integration. There were two distinct kinds of nodes – tool nodes and dataset nodes. Tool nodes were for the different tools that could be used to access information in the KB. The required information for a tool node included the startup information and the communication protocol. Dataset nodes represented a variety of files, dataset types, database queries (SQL) and research tools - all the contents of the KB as best as could be captured with one tool. Each dataset node included as much descriptive information as possible: proper name and version, location of the data on disk, dataset type, descriptive keyword information and the tool that should be used for access. A load date and the name of the author of the dataset node information were also included to keep track of when and who entered the information. The descriptive keyword information and the geographic extent of an integrated research product were taken from the documentation supplied by the Scientific Integrator for the product. There could be any number of keywords that describe an integrated research product. Most keywords were common words or phrases that could be used to describe the subject of the integrated research product. In addition, keywords could be a geographic name of the location covered by the integrated research product.

Displaying the data directly in Navigator itself only worked well for the Contextual Data integrated research products and Event Data integrated research products where we could define specific database queries and show the results. For most of the other products, configuration files were set up to display data using the appropriate tool for the integrated research products. Tools were configured to be launched from the KB Navigator using shell scripts that execute simple commands from the Navigator.

Descriptive information on the contents of the KB Navigator nodes and how to display or use the dataset was included in the README file started by the Validator for each integrated research product. If the integrated research product was not yet viewable with the KB Navigator, the information about where the dataset was located, why it was not viewable in the KB Navigator and the time frame for when it would be included was discussed in the README file. Every README file was added as a node in the KB Navigator so that a user would be able to easily see this information.

The final step in the integration process was to make sure that each integrated research product could be accessed and used through the KB Navigator. (Figure 6). This step ensured that the KB Navigator schema had been set up correctly, and again verified that no problems in accessing or using the integrated research product occurred during integration. This process also allowed the KB Integrator to assess the functional weaknesses of some of the products in the KB. It was difficult in some cases to determine exactly how products were intended to be used and what their format should be if they could not be run using one of the research tools, queried from the database or even highlighted in the document.

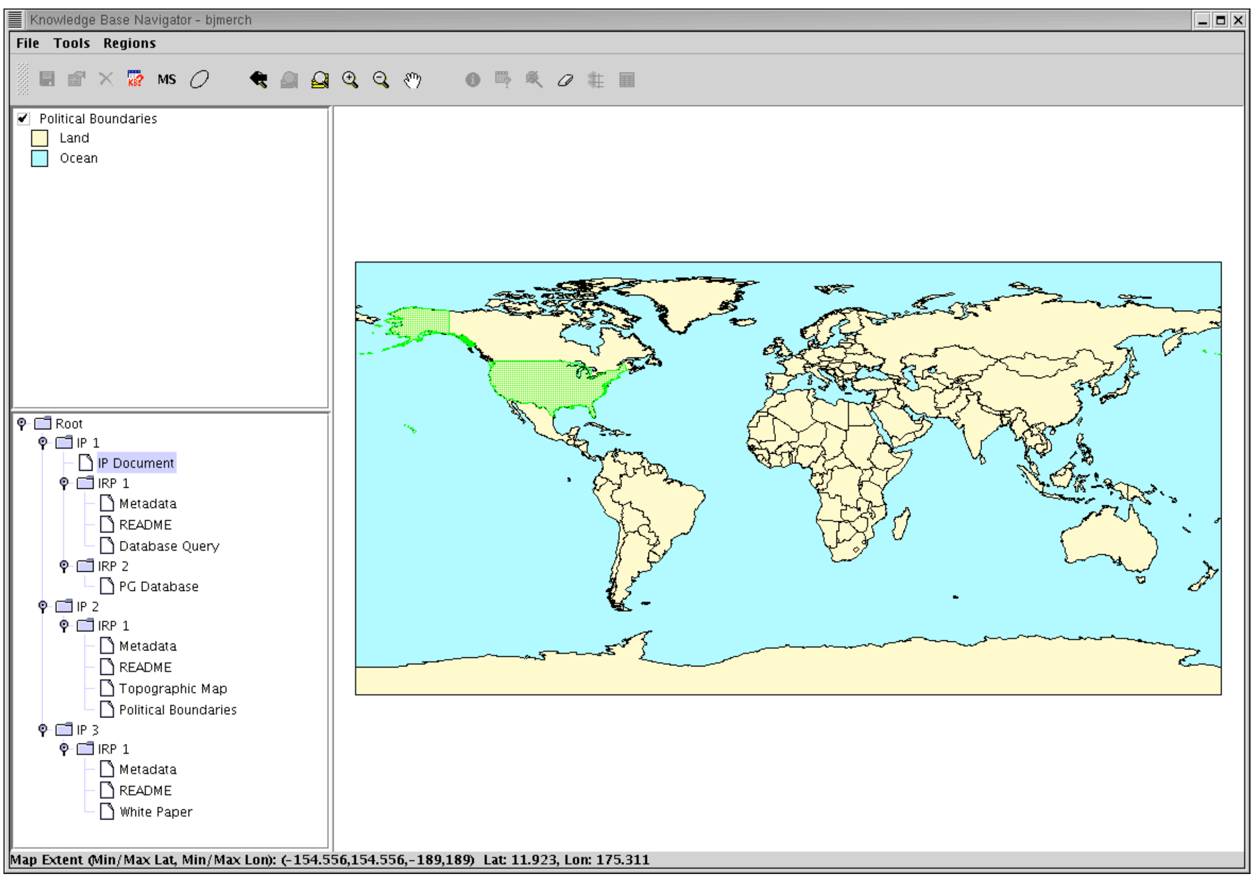

Figure 6. KB Navigator

Documentation

Documentation is also an important part of the KB. The main documents are the IP documents. The core of the IP document is the integrated research product descriptions written by the individual Scientific Integrators. The introduction section contains the overall purpose of the IP, summarizes the changes from previous versions, lists members of the working group that created the IP and provides an overview of the products in the IP. This section is written by the administrative points of contacts for the working group assigned to that IP. Each IP document was also assigned an editor to help put the introduction and integrated research product descriptions together and do final formatting and editing.

Scientific Integrators were required to get final descriptions for their integrated research product to the KB Integrator six weeks after the products themselves arrived at SNL. The editors came to SNL at that point and began compiling the integrated research product descriptions into the composite IP document. The editor also worked with administrative points of contact in getting the introduction written and into the document. The final documents were converted into PDF and placed in the Documents directory. All the IP documents were entered as dataset nodes in the KB Navigator so that users could easily access them.

CONCLUSIONS AND RECOMMENDATIONS

A well-defined process is needed when integrating products into the KB to make sure the result is a high-quality seamless product. Involving more people in the process of verifying and functionally validating integrated research products allowed the KB Integrator to take the time needed to really make sure a product could be accessed and used in the most appropriate manner. Running the Installation Test Procedures both before and after a product was integrated was important to let the KB Integrator know if any problems had been introduced due to integration. Finally, the use of the KB Navigator as a final test of the KB was inordinately valuable. SNL was very confident that we knew exactly what was in this release of the KB, and what the weaknesses were.

REFERENCES

Carr, D. (2002), National Nuclear Security Administration Knowledge Base Core Table Schema Document, Sandia National Laboratories, SAND2002-3005 (available at http://www.nemre.nn.doe.gov).

Gallegos, D. P., D. B. Carr, P. B. Herrington, J. M. Harris, C. L. Edwards, S. R. Taylor, J. J. Zucca, D. B. Harris, N. A. Wogman, D. N. Anderson and L. A. Casey (2002), The Integration Process for Incorporating Nuclear Explosion Monitoring Research Results into the National Nuclear Security Administration Knowledge Base, SAND2002-2772 (available at http://www.nemre.nn.doe.gov).

Hipp, J.R., R.W. Simons, L. Jensen, L. A. Lindsey and M. N. Chown (2002), The GNEM R&E Parametric Grid Data Software Suite: Tools for Creation, Access, Management, Viewing, and Export, *"Proceedings of the 24th Seismic Research Review: Nuclear Explosion Monitoring: Innovation and Integration",* 2002, Los Alamos National Laboratory *LA-UR-02-5048.*