BUILDING THE KNOWLEDGE BASE: TRANSITIONING SCIENCE FROM RESEARCH TO OPERATIONS

Christopher J. Young,[1] Preston B. Herrington,[1] J. Mark Harris,[1] C. L. Edwards,[2] Steven R. Taylor,[2]
John J. Zucca,[3] David B. Harris,[3] Dale N. Anderson,[4] Harry S. Miley,[4] Mark T. Woods,[5] and Leslie A. Casey[6]

Sandia National Laboratories,[1] Los Alamos National Laboratory,[2]
Lawrence Livermore National Laboratory,[3] Pacific Northwest National Laboratory,[4]
Air Force Technical Applications Center,[5] and National Nuclear Security Administration[6]

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

Contracts No. DE-AC04-94AL85000, W-7405-ENG-36, W-7405-ENG-48, DE-AC06-76RLO1830

ABSTRACT

Nuclear explosion monitoring depends upon the accurate, continuous, and immediate processing of a vast amount of data to support the needs of decision makers. The Air Force Technical Applications Center (AFTAC) operates highly automated data collection and processing systems to support this mission. This data processing system requires extensive knowledge (e.g., station calibration and tuning parameters) to effectively monitor the earth for signs of nuclear explosions.

The National Nuclear Security Administration (NNSA) Ground-Based Nuclear Explosion Monitoring Research & Engineering (GNEM R&E) program has developed the process, infrastructure, and tools needed to manage the movement of research results into the NNSA Knowledge Base for evaluation and potential use by AFTAC. When possible, information in the NNSA Knowledge Base is put into a form used by the operational software. Thus, AFTAC can readily incorporate this information into its testbed for validation and appropriate baseline inclusion into operational pipeline processing. The NNSA Knowledge Base also facilitates use of information in an analyst-driven, special event-processing mode, in some cases by providing software tools included in the NNSA Knowledge Base itself.

This paper describes the life of a scientific data product from its conception as a research effort and development at a research institution, through its inclusion in a NNSA laboratory information product to its integration into the NNSA Knowledge Base, where after delivery it is tested and evaluated by AFTAC for use in the automated processing system or by event analysis staff.

Introduction

In this paper we describe the process that NNSA has developed to integrate research products into an NNSA Knowledge Base release. This process produces a well-integrated, scientifically validated, fully documented NNSA Knowledge Base that is fully compatible with AFTAC's procedures and software and that will help them with their monitoring mission. It is essential not only that the NNSA Knowledge Base be effective, but also that it be easy to use. Despite the tremendous size and complexity of the NNSA Knowledge Base, we seek to deliver a product that is as close as possible to "plug-and-play" when it is delivered to AFTAC for evaluation.

AFTAC's Treaty Monitoring Mission

AFTAC is the US agency tasked with monitoring of the earth environment in the context of international nuclear treaties. AFTAC accomplishes this monitoring mission by acquiring various types of sensor data (seismic, atmospheric and space, hydroacoustic, infrasonic, and radionuclide) from the field environments, and then processing it to look for possible nuclear events.

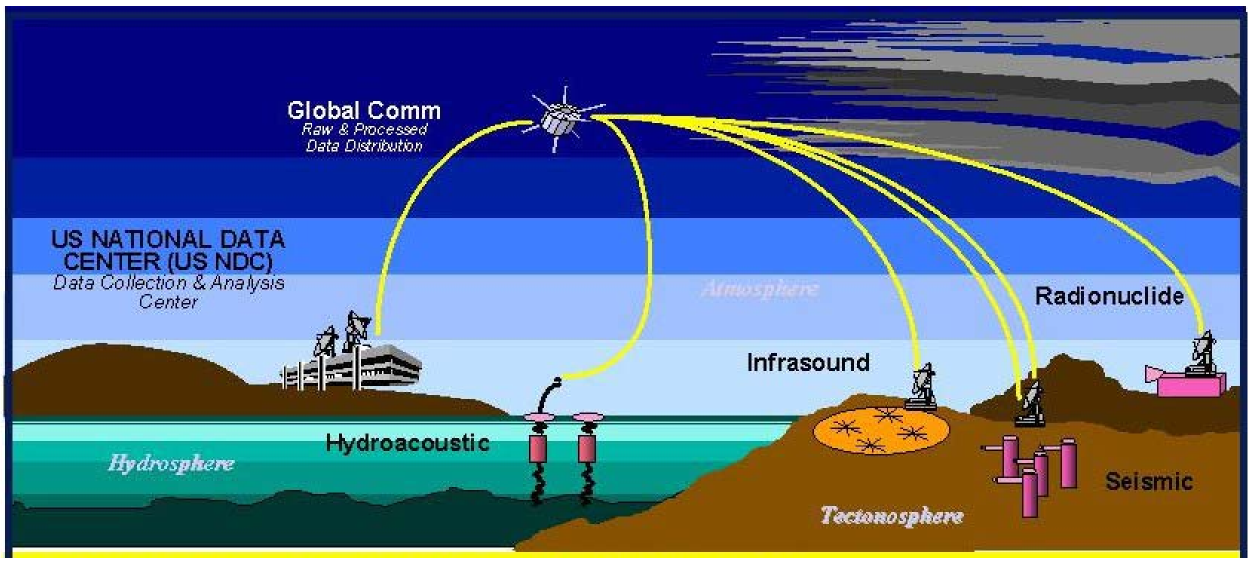

Figure 1. USNDC ground-based sensor data collection and transmission (from USNDC web site: http://www.tt.aftac.gov/overview.html).

The computer systems used to process the ground-based sensor data are in the United States National Data Center (USNDC) at AFTAC. The USNDC supports automated processing as well as interactive analyst review, and thus is the prime customer for any relevant monitoring research.

To better understand what types of research products might be of use for the USNDC, it is helpful to know a little about how the USNDC works. The core of the USNDC is an automatic data processing pipeline that feeds raw sensor data in at one end and produces a list of possible nuclear events at the other. To accomplish this, the data is processed through a series of steps that we briefly describe here in conjunction with Figure 2. Note that these steps hold for seismic, hydroacoustic, and infrasound data; processing for radionuclide data and data from satellite-based sensors is somewhat different.

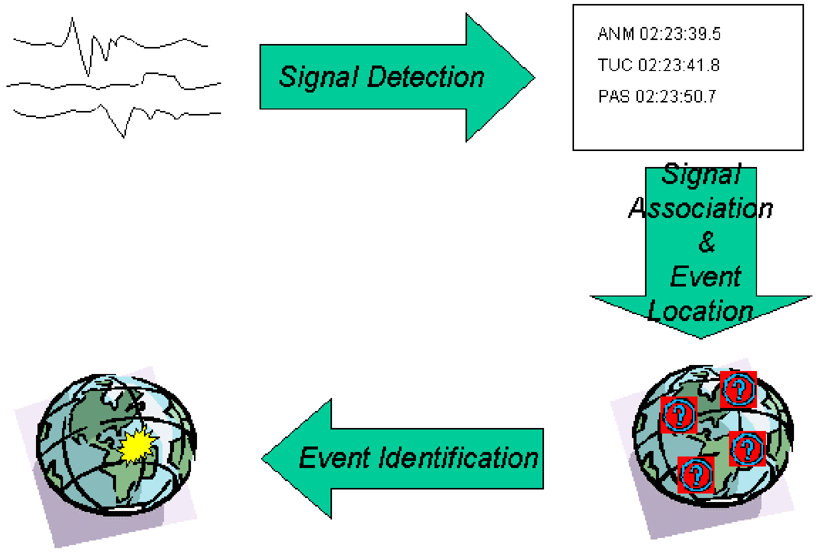

Figure 2. A simplified conceptual diagram of the USNDC pipeline processing tasks.

First, raw sensor data are processed to look for detections that might come from nuclear events. This analysis is a critical step because if no detections are found for an event, then regardless of the following steps, the event itself will not be found by the automatic processing. The next step is to group detections from the same event together. This grouping is known as association, and this step is closely related to the next step, event location, wherein the latitude, longitude, depth/elevation, and time of the event are determined. Once an event has been properly located, it can be categorized both by size (magnitude or yield) as well by event type (i.e., definite explosion, possible explosion, possible earthquake). The end result of processing data through these steps for a given time period is a list of well-located events with source types identified. Analysts then review these automatic event lists.

Every processing step is important, so improving any of them can have a significant impact on the quality of AFTAC's monitoring. Though the algorithms and procedures taken as a whole are diverse and complex, virtually all of them involve minimizing the difference between some observed quantity (i.e., raw or processed sensor data) and the corresponding predicted value (e.g., an arrival time of a seismic phase). Thus, to improve monitoring capability, one should seek either to improve the quality of the observations (e.g., more or better stations, better instrumentation, improved detection algorithms) or to improve the quality of the predictions (i.e., improved Earth models). NEM R&E includes both types of work, but the most voluminous contributions have focused on improving Earth models compiled in the NNSA Knowledge Base. Virtually any type of research product that can aid in modeling ground-based sensor data is of potential use for the NNSA Knowledge Base—i.e., 1-D travel time models, 3-D tomographic models, ground-truth event sets, topography maps, etc. There is no set list of types of products that do or do not belong in the NNSA Knowledge Base; if a research product can either directly or indirectly improve AFTAC's monitoring capability, then it should be considered for inclusion in the NNSA Knowledge Base.

The Role of the NNSA Knowledge Base

AFTAC requires an integrated, well-indexed, thoroughly validated NNSA Knowledge Base that is completely compatible with their methods and software. Conversely, most researchers' work is done in a developmental environment so they can focus on the important scientific issues, which are not generally dependent on specific data formats and operational algorithms. This gap between what AFTAC needs and what most researchers produce has historically resulted in a relatively small number of research products being utilized by AFTAC, despite the obvious potential relevance of many of them. AFTAC, as an operational user, does not have the resources to do the many things required to utilize these research products.

The NNSA GNEM R&E program has been very successful at bridging this gap. NNSA personnel are actively involved in monitoring research, so they are knowledgeable about the state of the art research. They are also familiar with the details of AFTAC's mission, so they can look for good matches between research products and operational use. In addition, they have an excellent support infrastructure they can draw on to help them transition products from research to operations. Besides geophysicists, NNSA personnel include experts on nuclear phenomenology, databases, data formats, data integration, metadata, etc. The NNSA Knowledge Base provides a mechanism to get research products used at AFTAC so that the research investment achieves the intended national security enhancement.

Integration of Research Products

The NNSA Knowledge Base integration process is documented in a report by Gallegos et al. (2002). A summary diagram is shown below.

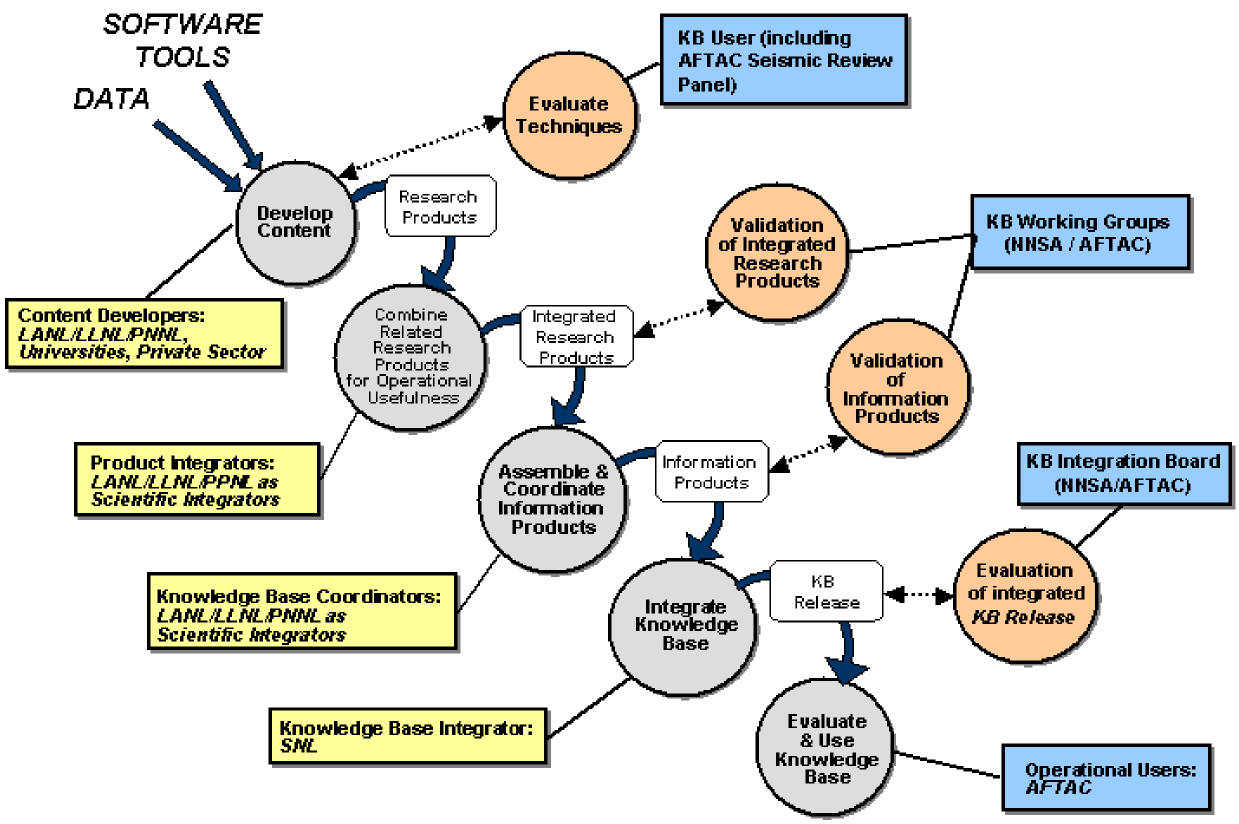

Figure 3. The NNSA Knowledge Base integration process

The process is complex and is outlined in the following 8 steps.

1. Develop Content

The first "step" in our process is really an ongoing process: scientists (*Content Developers*) constantly perform monitoring research and develop *Research Products* that have potential operational use at AFTAC. The necessary screening, reformatting, integration, and validation that enable these products to achieve their operational potential are performed in steps 2–8 of the NNSA Knowledge Base integration process.

2. Information Product Proposal and Approval

The next step in the process, the first real integration step, is to propose and plan the integration of *Research Products* into *Integrated Research Products* which are an aggregation and packaging of a suite of research products into an operationally useful form. *Integrated Research Products* then are further integrated into *Information Products* for the coming NNSA Knowledge Base release. An *Information Product* is a group of related *Integrated*

Research Products that address one of AFTAC's basic monitoring tasks (e.g., Detection, Association, and Location is one; Event Identification is another). The design of each of the Information Products is assigned to one of eight NNSA Knowledge Base *Working Groups* that are composed of researchers from the GNEM R&E labs and AFTAC. There is one Working Group per Information Product, and hence per AFTAC monitoring task. The *Integration Board*, comprised of NNSA and AFTAC program leaders evaluates each proposed *Integrated Research Product* for incorporation into the proposed NNSA Knowledge Base. The objective of the integration board is to assure that AFTAC's needs are addressed. They may reject an *Integrated Research Product* if it does not address an AFTAC need or suggest new products to fill unmet requirements.

3. Integrated Research Product Assembly

Once an *Integrated Research Product* proposal has been approved, the *Product Integrator* executes the plan to build it. The *Product Integrator* may solicit datasets or algorithms from the original contributor as part of the plan, and may refer the contributor to the NNSA Knowledge Base Contributor's Guide (Carr, 2002a) that guides him or her through the process of preparing datasets for transfer. For certain types of datasets, specific formats are required (e.g., Carr, 2002b), but often the formatting is done by the *Product Integrator* after receiving the *Research Product* from a contributor. The *Product Integrator* may ask the contributor for metadata, which help provide an audit trail for construction of the dataset or algorithm.

As part of product assembly, the *Product Integrator* will reformat data or recode algorithms to be compatible with AFTAC software. Extensive testing of the finished product follows to assure proper operational use. The *Product Integrator* creates extensive *Integrated Research Product* documentation (which includes any metadata), product installation procedures and tests, following NNSA templates.

4. Information Product Integration and Validation

By specified deadlines for each release, the *Integrated Research Products* (content and documentation) must be presented by the Scientific Integrator to Sandia National Laboratories (SNL), as the NNSA Knowledge Base Integrator, for integration in conjunction with the appropriate Working Groups into *Information Products*. SNL validators check product contents against documentation, and attempt to run installation tests. The validators iterate with *Product Integrators* to resolve any problems encountered.

In parallel with the integration and testing of the content of the *Information Product* at SNL, an *Information Product* document is created by one of the *Information Product Editors* based at Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory (LLNL) from component *Integrated Research Product Documents* sent to the *Editors* by *Knowledge Base Coordinators* at each laboratory. When each *Information Product* document is complete, it is sent to the *Knowledge Base Integrator* for assembly into comprehensive documentation for the entire NNSA Knowledge Base.

5. NNSA Knowledge Base Integration and Testing

Once the *Integrated Research Product* components of the *Information Products* have been put in place and unit level tested, the proposed NNSA Knowledge Base is assembled. A graphical interface to the proposed NNSA Knowledge Base is built using a new NNSA tool known as the Knowledge Base Navigator. The Navigator provides a simple "point-and-click" style interface that allows users rapid and intuitive access to the vast holdings of the Knowledge Base.

The tool combines a hierarchical file browser with a map that can be used to show the spatial extents for *Integrated Research Products* or to find *Integrated Research Products* by specifying a spatial location (e.g., the border of a given country). The Navigator also can launch all of the software tools delivered with the NNSA Knowledge Base. Working Group correlated *Information Products* are organized as folders within the Navigator interface; constituent *Integrated Research Products* are found in subfolders. Within the *Integrated Research Product* subfolders are additional folders that hold product components and automatically launch the installation tests for the product. Folder contents can be associated with keywords such as "group velocity model" and "phase match filtering" so that they can be found using Navigator's keyword search capability. Contents also can be linked with a geographic polygonal region showing the area of coverage, enabling the product to be found by a geographic search.

6. Integration Board Approval

The *Integration Board* reviews individual *Integrated Research Products*, and assesses the quality of the proposed NNSA Knowledge Base as a whole, aided by the Navigator and the now complete documentation in the accompanying *Information Product* document. This review can result in deferral of products to later releases to allow more development or redirection of products based primarily on optimizing operational usefulness. The *Board* assures that products pass all integration crosschecks, that all content has been delivered, and that validation testing has been passed. It also verifies proper integration of product documentation.

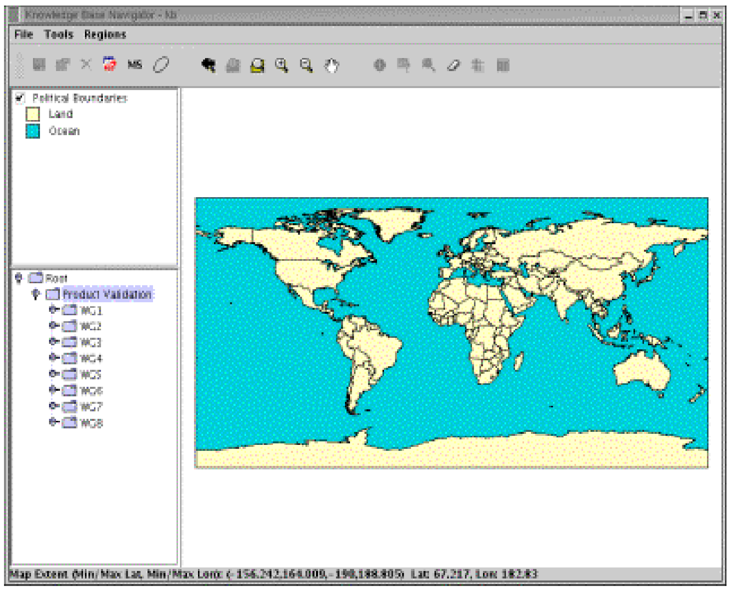

Figure 4. A fictional KB as presented through the KB Navigator

7. NNSA Knowledge Base Delivery and Demonstrations

The approved NNSA Knowledge Base is delivered to AFTAC, which involves more than a simple transfer of files. Some re-configuration must be done on-site to account for any system differences between the NNSA Knowledge Base development test area at SNL and the AFTAC target system. Products are re-tested to assure that nothing has been broken during delivery. The fully tested NNSA Knowledge Base delivery is demonstrated to AFTAC by teams of NNSA researchers who travel to AFTAC, describe the use of the products and provide hands-on training.

8. AFTAC Evaluation of the NNSA Knowledge Base

AFTAC personnel critically evaluate the NNSA Knowledge Base. Generally NNSA Knowledge Base calibration products and tools are evaluated on AFTAC data to test effectiveness for various monitoring tasks. Performance is the most important criterion, but organization, ease of use, and completeness of metadata are also factors. Detailed feedback for each product is provided to NNSA *Product Integrators*, which guide product updates and new product development. Following evaluation, AFTAC chooses suitable parts of the NNSA Knowledge Base for implementation and invoke their internal configuration control procedures.

List of Data Types Integrated

Table 1 lists data types that have been, or are expected to be, integrated by the NNSA program. The table lists also the type of scientific integration performed (second column) and the end use or product developed (third column). The table makes clear that the NNSA program seeks a very wide range of data, models, and algorithms for integration into the NNSA Knowledge Base. We make no attempt to describe the details of this table here, but present two representative examples of integration.

Table 1. Data types integrated by the NNSA program, integration performed, and end products

Data Type	Integration Performed	End Uses / Products
Catalogs, bulletins • Event descriptions • Picks • Aftershock studies	Reconciliation of all catalogs into a unified master catalog Ground-truth event collection and validation Travel-time tomography and other model construction Kriged travel-time correction surface generation	Driver for event waveform segment extraction Travel-time model validation Velocity models Phase travel-time correction surfaces
Waveform data, from: • Major providers: AFTAC, IRIS, USGS • PASSCAL deployments • Joint BAA/ROA deployments • Local network data • Other sources	Event waveform segment extraction – waveform database construction Phase arrival time and amplitude measurement integration into unified catalog Full-waveform modeling to estimate source parameters and velocity models Source phenomenology studies Coda shape calibration	Phase amplitude correction surfaces MDAC amplitude correction models Source models Discriminants Coda magnitude calibrations Base velocity models
Source ground-truth information • Earthquake mechanisms, depths, moments • Explosion yields, shot configuration data, emplacement data	Coda parameter estimation MDAC parameter estimation Discriminant calibration Source phenomenology studies	Coda magnitude calibration and yield estimation MDAC amplitude correction models Explosion models Discriminants
Geophysical information • Velocity models • Surface wave slowness maps • Attenuation maps • Geologic maps • Contextual data	Validation on AFTAC and NNSA test data Travel-time ray tracing MDAC parameter estimation Unified geophysical model construction	Base geophysical models Travel-time models MDAC amplitude correction models Contextual database for analysts Ms phase-match calibrations
Derived Measurements • Surface wave dispersion	Surface wave dispersion and attenuation tomography	Surface wave dispersion maps and models Moment tensor estimation Ms phase-match calibrations
Algorithms • Signal processing • Array • Location and discrimination • Error propagation • Inversion	Validation on AFTAC and NNSA test data Assessment for calibration potential Recoding to use AFTAC / KB pipeline software libraries	Pipeline detection, global association, location and identification functions Analyst event review/reanalysis Calibration methods Error assessment

Representative Integration Examples

Our first example is the construction of a surface wave dispersion map useful in phase-match filtering to estimate Ms and potentially full-waveform moment tensor inversions. A very high-quality, five-degree, global surface wave model is available from Stevens and McLaughlin (1997), which the NNSA labs are using as a baseline (Figure 5, where the top panel shows the 20-second Rayleigh velocity map). We are developing complementary, two-degree, high-resolution regional maps to be embedded into the Stevens-McLaughlin background model. Results of this embedding process are illustrated in the middle and lower panels of Figure 5. The middle panel shows the same model overlayed by an NNSA-developed, high-resolution Western Eurasia and North Africa (WENA) tomography model. Slow velocities from sedimentary basins like the Caspian Sea and Persian Gulf are more clearly delineated in the regional model. Finally, the bottom panel shows the blending of the WENA models with the Stevens and McLaughlin model, providing global coverage.

To build the WENA high-resolution regional velocity model, we systematically measured Rayleigh and Love wave group velocities (Pasyanos et al., 2001) and integrated measurements from a number of other research groups. In the WENA region we examined more than 25,000 seismograms and made more than 15,500 Rayleigh wave measurements and 10,000 Love wave measurements at periods from 8–150 seconds. We exchanged group velocity curves with other research groups and incorporated more than 3,000 path measurements from the University of Colorado (Ritzwoller, written communication) and SAIC/Maxwell (Stevens, written communication).

Our second example illustrates the extension of the 3-degree Eurasian coda Q map for 1 Hz Lg of Mitchell *et al.* (1997) with a 2-degree inset attenuation tomography map (Figure 6, white outline). The high-resolution regional map was developed using a set of 1,651 events recorded at 12 stations. For central Asian events, this high-resolution map provides a variance reduction of 27% over the prior model of Mitchell *et al.,* (1997) and 67% over an initial constant Q model of 418 (the average of the Mitchell *et al.,* 1997 model).

We have been investigating the utility of Bayesian methods as a way to incorporate and refine tomographic maps as part of the integration process. By using a Bayesian approach, tomographic maps provided to the NNSA laboratories can be used as prior background models in order to produce high-resolution refined maps in critical areas. The resulting refined tomographic models smoothly blend into the prior background models. Moreover, the uncertainties are well established in a Bayesian framework. We have been using Bayesian tomography for regional phase attenuation models (Taylor *et al.,* 2003), surface wave slowness models (Maceira and Taylor, 2003), and surface wave attenuation models (Yang *et al.,* 2003).

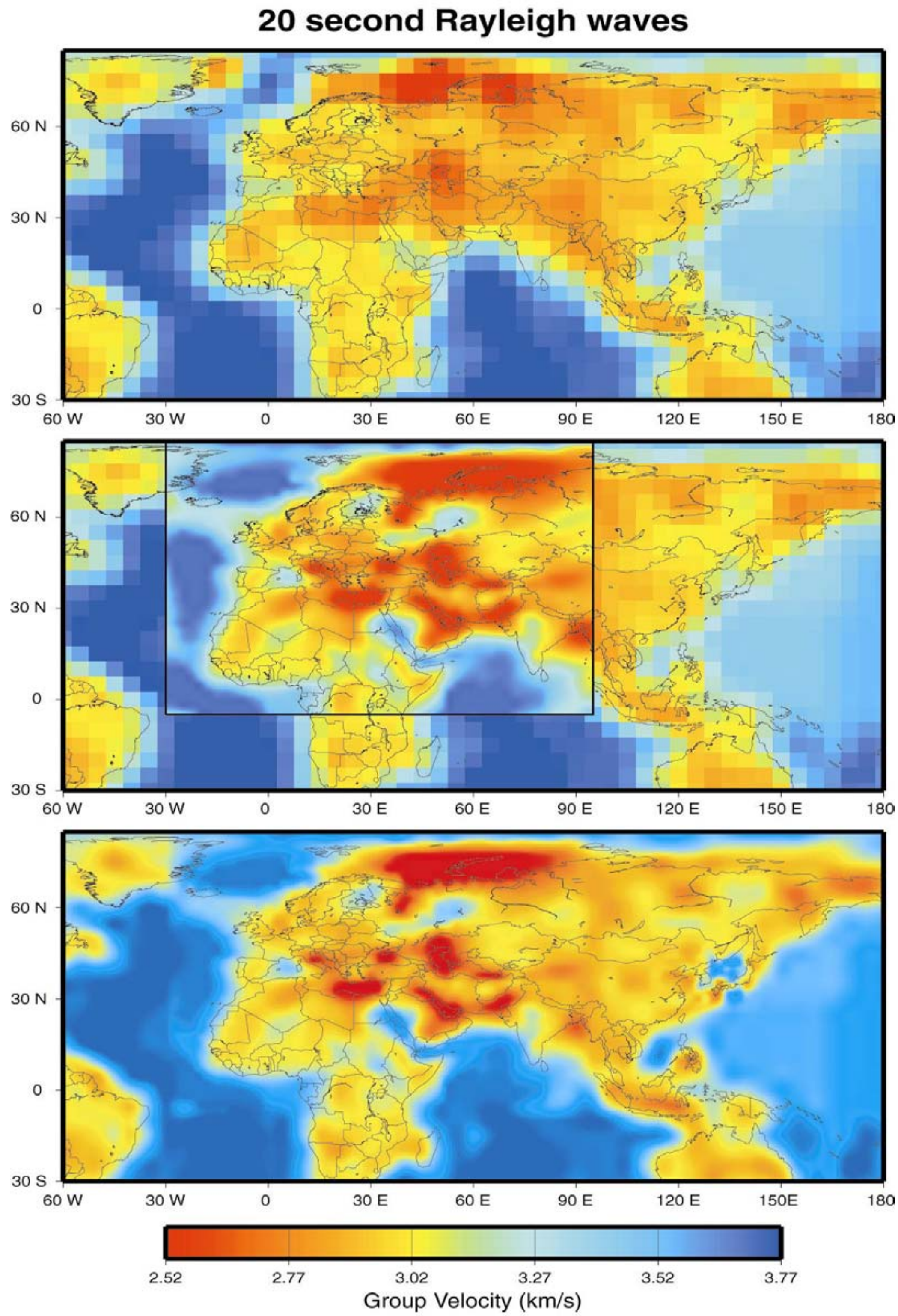

Figure 5 Example of blending two-degree regional surface wave tomography maps into a five-degree global baseline map. Original 20-second Rayleigh wave group velocity map (top) of Stevens and McLaughlin (1997). High-resolution inserts superimposed (middle). Blended map with global coverage (bottom). The five-degree background map has been interpolated to the higher-resolution grid.

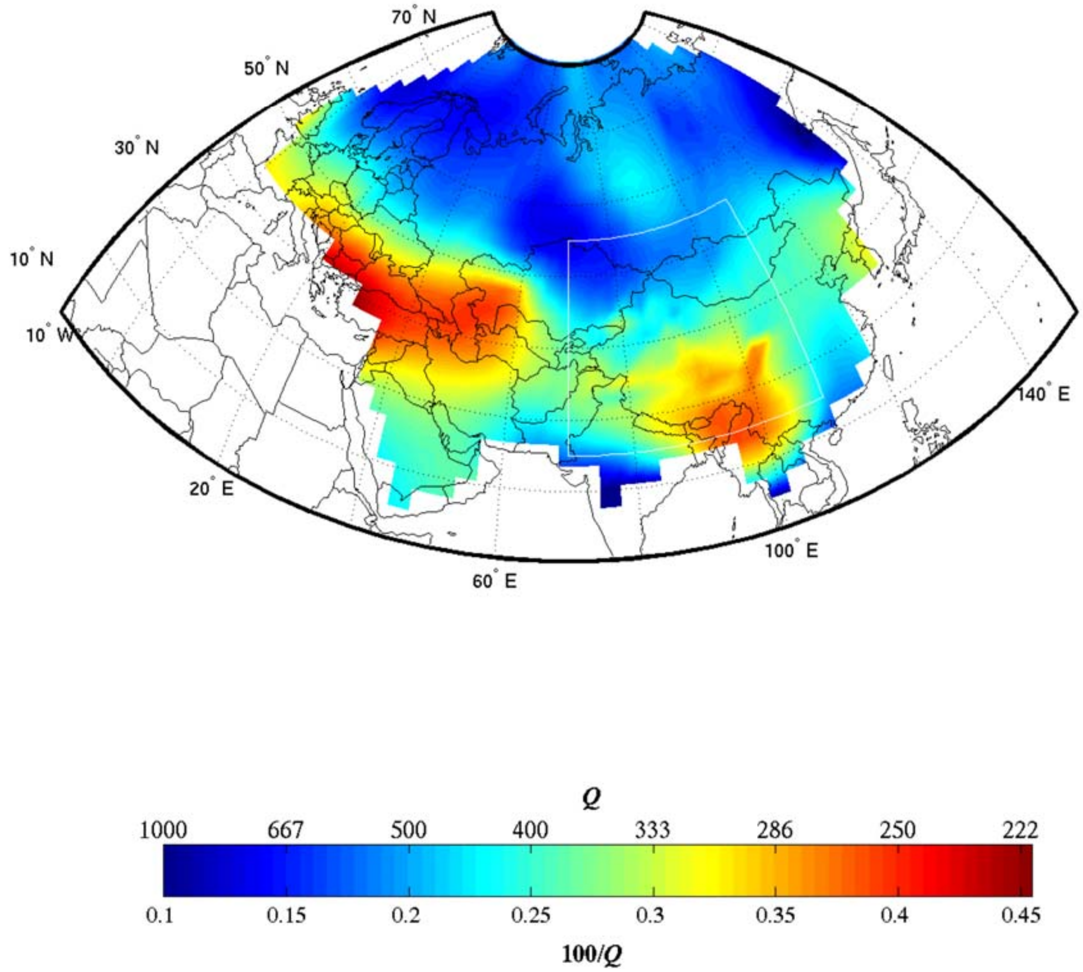

Figure 6. Two-degree attenuation tomography map for 1 Hz Lg of Figure 2 (white outline) blending into 3 degree Lg coda Q map of Mitchell et al. (1997).

SUMMARY

The valuable *Research Products* from relevant nuclear explosion monitoring research (largely represented at this Seismic Research Review) must be put into an operationally useful form to effectively improve national security. The NNSA GNEM R&E program has developed the process, infrastructure, and tools needed to manage the movement of nuclear monitoring research results into the NNSA Knowledge Base for evaluation and use by AFTAC for their nuclear treaty monitoring mission. The NNSA Knowledge Base integration process has greatly increased the number of research products that reach operational usefulness at AFTAC. The process exploits unique NNSA resources (personnel and support infrastructure) to bridge the gap between the disparate worlds of nuclear monitoring research and nuclear monitoring operations, thereby greatly benefiting both AFTAC and the research community by making optimum use of government resources.

REFERENCES

Carr, D. B. (2002a), National Nuclear Administration Security Knowledge Base Contributor's Guide, Sandia National Laboratories Report SAND2002-2771 (Revised, available at https://www.nemre.nnsa.doe.gov/cgi-bin/prod/nemre/index.cgi?Page=Knowledge+Base).

Carr, D. B. (2002b), National Nuclear Security Administration Knowledge Base Core Table Schema Document, Sandia National Laboratories Report SAND2002-3055 (available at https://www.nemre.nnsa.doe.gov/cgi-bin/prod/nemre/index.cgi?Page=Knowledge+Base).

Carr, D. B (2003), Assembling the NNSA Knowledge Base, 25th Seismic Research Review - *Nuclear Explosion Monitoring: Building the Knowledge Base*.

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, Sandia National Laboratories Report SAND2002-2772 (available at https://www.nemre.nnsa.doe.gov/cgi-bin/prod/nemre/index.cgi?Page=Knowledge+Base).

Maceira, M. and S.R. Taylor (2003), Short-period surface-wave tomography in central Asia and its application to seismic discrimination, *Seism. Res. Lett*., 74, 212.

Mitchell, B.J., Y. Pan, J. Xie, and L. Cong (1997), Lg coda Q variation across Eurasia and its relation to crustal evolution, *J. Geophys. Res.,* 102, 22,767-22,779.

Pasyanos, M. E., W. R. Walter and S. E. Hazler (2001). A surface wave dispersion study of the Middle East and North Africa for monitoring the comprehensive nuclear-test-ban treaty, *Pure. Appl. Geophys*., 158, 1145-1474.

Stevens, J.L., and K.L. McLaughlin (1997), Improved methods for regionalized surface wave analysis, Phillips Laboratory Report PL-TR-97-2135, 26 pp.

Taylor, S.R., X. Yang, and W. S. Phillips (2003), Bayesian Lg attenuation tomography applied to eastern Asia, *Bull. Seism. Soc. Am.*, 93 795-803.

Yang, X., H.J. Patton, and S.R. Taylor (2003), 20-sec Rayleigh-wave attenuation tomography in central Asia, *Seism. Res. Lett.,* 74, 213.