Collaborative Development of a US/Russia Safety Information Center Database

M. A. Dvorack
Assessment Technologies Department, Sandia National Laboratories, P. O. Box 5800, Albuquerque, New Mexico 87185-0405 USA

R. E. Smith
FSU Project, Sandia National Laboratories, P. O. Box 5800, Albuquerque, New Mexico 87185-0491 USA

V. N. Ananiychuk
Russian Federal Nuclear Center, All-Russian Research Institute of Technical Physics, Chelyabinsk-70, P. O. Box 245, Snezhinsk, Chelyabinsk Region 454070 Russia

N. A. Volkova
Russian Federal Nuclear Center, All-Russian Research Institute of Experimental Physics, Arzamas-16, Sarov, Nizhni Novgorod Region 607200 Russia

ABSTRACT: One of the major outgrowths resulting from the collapse of the former Soviet Union (FSU) has been an increase in technical information exchange and dialogue between the Russian and American nuclear weapons laboratories. One area of such discussions is concerned with the safety of low probability, high consequence systems and operations. In order to further the understanding between the respective institutes in this important area, a collaborative effort has been established between Sandia National Laboratories and the two premier Russian nuclear weapons laboratories, Arzamas-16 and Chelyabinsk-70, in which a common database has been developed which contains safety information provided by all three laboratories. More than 1200 documents have been placed by the three institutes into this database. This paper describes the details of this database, including the types of safety information being stored.

HISTORICAL BACKGROUND

1. Introduction

One of the major outgrowths resulting from the collapse of the former Soviet Union (FSU) has been an increase in technical information exchange and dialogue between the Russian and American nuclear weapons laboratories. One area of such discussions is concerned with the safety of low probability, high consequence systems and operations. In order to further the understanding between the respective nuclear weapons laboratories in this important area, a collaborative effort has been established between Sandia National Laboratories and the two premier Russian nuclear weapons laboratories, Arzamas-16 and Chelyabinsk-70, in which a common database has been developed which contains safety information provided by all three laboratories. The goal of this collaborative activity is to develop a user-friendly database which will contain important reference information in the area of safety for use by appropriate government agencies in both Russia and the United States. The ultimate goal is to make this information available to the international safety community.

As implied above, both the United States and Russia are concerned about the risks and consequences of low probability, high consequence systems. Furthermore, the scientists and engineers from both countries who are involved with such systems are eager to draw on the experiences and, especially, the lessons learned from their colleagues. Hence, the database provides an opportunity for Russian and American engineers to learn about each other's experiences, as well as to become more familiar with each other's capabilities regarding various issues pertaining to safety. This knowledge may thus provide the Russian and American engineers useful information which they may adapt to their own use for addressing safety issues regarding the systems on which they are working.

We expect to develop a Web accessible, user friendly data access and retrieval information system that is fully functional in both English and Russian. To date, the database system is non Web-based and accomplishes this goal for the English text documents and essentially stores the Russian documents for image retrieval.

This system is the result of several collaborative discussions between the Russian and American engineers at the three institutes. Funding for the two Russian institutes in their efforts to help develop the database was provided by Sandia under contract AK-5005 for Chelyabinsk-70 and contract AK-5006 for Arzamas-16, both contracts entitled "Development of a Surety Information Center Database for High Consequence Engineered Systems."

BRIEF DESCRIPTION OF THE INSTITUTES

Since many people are not familiar with the nuclear weapons laboratories of the United States and Russia, a brief summary describing each institute is provided in the sections below.

1. Sandia National Laboratories

Sandia National Laboratories is a United States

DISCLAIMER

Department of Energy (DOE) multi-program laboratory operated by Sandia Corporation, a wholly-owned subsidiary of Lockheed Martin Corporation. Sandia originally was part of Los Alamos National Laboratory until 1949, when President Truman asked AT&T Corporation to manage Sandia. Eventually, management of the laboratories was awarded to Martin Marietta Corporation (now Lockheed Martin). Los Alamos and Lawrence Livermore are the physics laboratories while Sandia is the engineering design laboratory. Sandia has two main facilities, one at Albuquerque, New Mexico, and the other at Livermore, California. Test sites are located in Nevada and Hawaii.

Sandia's main responsibilities include providing engineering design for all non-nuclear components of American nuclear weapons, and is also responsible for maintaining the viability and safety of the American nuclear arsenal. In addition, Sandia is also involved in non-nuclear weapon research, such as robotics, microelectronics, nuclear reactors, computational and information systems, engineered processes and materials, and engineering sciences.

2. Arzamas-16

Arzamas-16, also known as the All-Russian Institute of Experimental Physics, was established in 1946 near Sarov, Russia. Arzamas-16 was the first of the Russian Federal Nuclear Centers where the manufacture of the first Soviet nuclear weapon components and associated laboratory testing was done.

Arzamas-16 is a complete nuclear weapon design organization involving the design of nuclear weapon charges and warheads, as well as various research and testing programs, including nuclear power safety, disarmament and nonproliferation activities, and industrial technologies applications.

3. Chelyabinsk-70

Chelyabinsk-70, also known as the All-Russian Institute of Technical Physics, is a Russian Federal Nuclear Center that was established in 1954, near Snezhinsk, Russia. In addition to having been involved in nuclear weapon development work, Chelyabinsk-70 is currently pursuing research activities such as the modeling of nuclear explosions and technology transfer.

DATABASE DESCRIPTION

The US/Russia safety database provides a means of capturing historic information and assisting potential users in the analysis of low probability, high consequence engineered systems. In addition to safety issues, the database also includes information on such topics as reliability and human factors; hence, because of these extra topics of information, the database is also referred to as a "surety" database. In fact, the database covers a wide range of unclassified topics of mutual interest to all three institutes, such as fire and explosion safety, risk assessment of engineered systems, radioactive disposal, explosion physics, and human factors in engineering. To date,

34	Litvinov B.V., Novikov G.A., Schukin V.A.	Safety of a nuclear arms system from a position of the theory of man-machine systems and industrial ecology	1995	Safety of labour in a industry, No. 5, p.33-35
35	Mazur I.A., Moldavanov O.I.	Introduction in engineering ecology	1989	M., Nauka
36	Mazurov V.D.	A method of committees in problems of optimization and classification	1990	M., Nauka
37	Moiseev N.N. et al	The person and biosphere. Experience of system analysis and experiments with models	1985	M., Nauka
38	Molodtsov A.	Sociological modeling as a method of effective management of staff	1991	Newsletter "Personal". All-Union Correspondence university of management of staff, Kiev, No. 2, p. 75-88
39	Muromtsev Yu. L.	Absence of accidents and diagnostics in chemical production	1990	M., Khimiya
40	Nikolaev V.I., Bruk V.M.	System engineering: methods and applications	1985	Leningrad, Mashinostroyenie
41	Novikov G.A., Schukin V.A., Tyurin G.A.	Principles of functioning of a national nuclear weapon maintenance system	1993	Report of a Center of the public information on nuclear energy, No. 11, p.52-55

Table 1. Sample listing of documents selected by the Russian institutes for inclusion in the US/Russia safety database .

more than 1200 documents have been archived by the three institutes since this collaborative effort began.

The current database has both image and full text storage and retrieval capabilities in English, and image capability in Russian. Currently, use of the database has been restricted to those scientists and engineers at the three institutes. However, the database will be made accessible to all appropriate agencies and organizations through future negotiated agreements as appropriate.

SYSTEM DESCRIPTION

1. System Hardware

Figure 1 illustrates the hardware that is used to create, maintain, and update the safety database. The system hardware shown in this figure was provided to the two Russian institutes by Sandia under a separate contract so that all three facilities would have common hardware and software configurations.

2. System Software

The software working environment for the database work at the two Russian institutes is Microsoft Windows for Workgroups 3.11, while Sandia uses Microsoft Windows 95. [1,2]. The primary software package used in creating the database at the Russian institutes is LaserFiche for Windows Version 3.0d, while Version 4.1g is used at Sandia [3]. LaserFiche allows the user to import and export files; scan individual documents or batches of documents; use optical character recognition (OCR) for creating searchable text from document images; search and retrieve documents; edit documents; and archive the docu-

No.	Fields	Content Field	Notes
1	2	3	4
1	NSIC#	Document number given by the Institute	Contains the Institute code and serial number
2	TITLE/SUBJECT	Document title or a subject: title of a conference or exhibition where the document was published	
3	SHORT TITLE	Short document designation	For instance, Contract AK-5006
4	AUTHOR	Author(s) of the document. Name and initials	
5	SHORT ABSTRACT	Document short abstracts	
6	KEY WORDS	Key words	6 words of no more than 25 characters length, not used in and template field
7	PAGE	Document page number	
8	TYPE	Document type	
	ARTICLE, AGENDA, BRIEFING, BULLETIN, CLIPPING, CHART, DRAWING, FAX/TWX, INDEX, INSTRUCTION, LAYOUT, LETTER, LIST, MANUAL, MAP, MEMO, MINUTES, ORDERS, PHOTO, REPORT, ROSTER, TEST_DATA,VIEWGRAPH, VIDEO, ELECTRONIC FILE		
9	DEPOSITORY	Original depository	
10	CORE_DOC	Document: quidance, policy, defining, or legal	One character "Y" or "N"
11	RELEASE_CODE	Publication permitting code 0 - no restriction (including open publication) 1 - Governmental agencies 2 - Military Departments (including Ministry of Atomic Energy and Department of Defense) 3 - Department of Energy or Ministry	
12	DOCUMENT_DATE	Date of document entering the DB	
13	TITLE_ISSUE	Title of publisher of a book, conference proceeding, etc.	

Table 2. English language template used at Arzamas-16 for storing documents in the US/Russia safety database. The acronym NSIC in line 1 refers to Sandia's Nuclear Safety Information Center.

ments in various folders. Because LaserFiche had been demonstrated to be a versatile, easy to use software package at Sandia, it was chosen to be the tool for storing documents and updating the database. For example, LaserFiche provides the capability to easily capture a wide variety of documents. Documents can be scanned individually or batch processed. A centralized filing system is also provided by this software package, allowing the user to create folders for the documents, as in a traditional filing system. LaserFiche also provides great flexibility in retrieving documents, allowing the user to retrieve documents by any word or phrase in the document's content, as well as by providing for the use of boolean operators to aid in the searching. A "fuzzy search" option is also provided [4].

LaserFiche for Windows Version 3.0d or 4.1g requires the computer system to have an operating system that is Microsoft Windows Version 3.11 or Windows 95, 16 MB of memory, a Windows,-compatible monitor display, and a 486 minimum central processing unit (CPU).

While LaserFiche for Windows Version 3.0d or 4.1g does an excellent job of performing full text search and other manipulations in English, it is unable to recognize Russian (i.e., Cyrillic) text. In other words, while it can easily handle both images and text written in English, it is unable to do so in Russian. Hence, images of Russian documents that were scanned could not be converted into searchable text using OCR. Thus, the search capabilities using this software are limited to English only.

Another limitation of the system is the lack of a built-in suitable Russian/English translation software package. Currently, users who are not familiar with one of the two languages must either rely on their own choice of translation software or have it converted by a translator. At present, the machine translator that Sandia is using is called Stylus for Windows Version 2.21 [5] which was developed in Russia. This software package has the capability of translating text from Russian into English and vice versa. The machine translation provided by Stylus provides an initial rough translation that is useful for determining the potential interest of the document in question. If the user needs to have a detailed understanding of the document, it is strongly recommended that a professional translator who has knowledge of the subject matter provide a final translation. The translated document should then be added to the database.

METHOD OF DATABASE ENTRY

Document Selection and Storage

Placement of documents into the database can be illustrated by describing the process used at the Russian institutes [6]. The process begins by performing a search in Russian science and technology journals, followed by the selection of those documents deemed suitable for entry into the database. An example of the results of this selection process is illustrated in Table 1 [2]. A template containing key information describing each of the selected documents is then created in both Russian and English. The documents are then scanned onto the internal 4 gigabyte hard drive. The documents, texts and images, are then exported to a briefcase file which is stored onto a Bernoulli® 230 MB disk. This disk is sent to Sandia for uploading into the master system. Sandia updates the Russian LaserFiche system using read-only compact disks (CD-ROMs). Copies of the CD-ROMS containing the data in the master database are sent from Sandia to Arzamas-16 and Chelyabinsk-70. This process is illustrated in Figure 2.

TEMPLATE DESCRIPTION

For each document that is placed into the database, a corresponding template is filled out in English and in Russian. The template contains key information regarding the document such as title, author(s), subject matter, number of pages, and key words. The English language template can also be used to search for a document [4]. The contents of such a template are illustrated in Table 2 and an example is provided in Table 3 [7].

USE OF THE DATABASE

For the English-speaking user, the best case scenario is being able to do a full text search on both the templates and documents with respect to the topic of interest, resulting in the identification of documents that are written in English. The worst case scenario for this user, on the other hand, would be the identification of documents that are written only in Russian. While there are templates for these documents that are written in English, the documents themselves would be in the form of images only. At this point, the reader would have to make use of optical character recognition software that can transform Cyrillic characters into text. At Sandia, the software that is used for creating text from Russian document images is called FineReader Version 2.0 [8]. This software, purchased in Russia, performs optical character recognition on both Russian and English document images. Once the Russian document image has been converted to text, the user can then use a machine translation program like Stylus. Machine translations currently are useful for translating a sufficient amount of text in order to get a general, but limited, understanding of the original material. If the machine-translated text indicates to the user that this is indeed a useful document, then a translation per-

formed by a qualified translator can be obtained and inserted into the database.

For the Russian-speaking user, a search of the database for a particular topic would necessitate initially translating the terms to search for into English, and then performing a search. Once the English language template is identified, the corresponding document (text, if English or image, if Russian) could then be retrieved. If the search results in the identification of Russian document images, the user can use the images directly. The user could also use additional software like FineReader in order to create Russian text from the images. With regard to English documents resulting from a search, the user must either rely on a qualified translator and/or utilize machine software like Stylus in order to obtain a Russian translation.

RESULTS TO DATE

As stated earlier, more than 1200 documents have been placed into the database by the three institutes. Maintenance and updating of the database, along with periodic releases of updated versions of this database, are provided to the two Russian institutes by Sandia.

To date, the database has provided useful insights for the engineers and scientists at all three institutes. For example, Russian documents placed into the database have enabled Sandia engineers to begin to understand how risk is perceived and addressed at the two Russian institutes. In addition, terminology differences and jargon, which can cause confusion when conducting technical discussions between American and Russian scientists and engineers, are also being identified [9]. Furthermore, both Russian and American scientists and engineers now have access to reports that were previously unavailable to them.

FUTURE WORK

Recently, discussions between the three institutes have been held at Sandia regarding the future development of this database. These discussions centered around the two main aspects of any database: (1) the information contained in the database, and (2) the ease of retrieval. Hence, these deliberations centered upon five specific areas, outlined below. The first area addresses the continuation of providing documents to the database, while the remaining four areas are addressed in order to improve system capabilities:

1. *Continuation of adding documents to the database and improvement of data in the existing documents.*

Adding additional documents is considered to be the first priority in maintaining the current US/Russian safety database. Negotiations are currently underway between the three institutes to incorporate another 200 documents minimum into this database.

2. *Development of Web capability at Arzamas-16 and Chelyabinsk-70 for the purpose of putting the database onto the Internet.*

Internet access of the database will achieve two main goals: (a) provide easy access to authorized users, both current and future, and (b) make the database available to a wide variety of potential customers.

No.	Title of the Field	Contents
1	NSIC#	000005EF
2	TITLE/SUBJECT	The Method of Active Remote Detection of Nuclear Ammunition
3	SHORT TITLE	
4	AUTHOR	Zalyubovsky A.A., Lomako A.A., Morgun O.N., Cherny V.V.
5	SHORT ABSTRACT	Method of Active Indestruction Remote Detection of Nuclear Weapons at Mobile or Remote Facilities with the Use of an Intermediate Energy Neutron Pulse Generator
6	KEY WORDS	expresness, gamma-response, ion accelerato
7	PAGE	4
8	TYPE	article
9	TITLE_ISSUE	Journal of the Atomic Energy, 1993, v. 74, #6, p.497
10	CORE_DOC	N
11	RELEASE_CODE	0
12	DOCUMENT_DATE	01.12.95

Table 3. An example of an English language template created at Arzamas-16 for a Russian document, using the outline of Table 3.

Currently, those who wish to use the database must contact the appropriate individuals at the three institutes in order to view the data. As presently set up, only Sandia, Arzamas-16, and Chelyabinsk-70 have the proper system hardware/software configuration needed to access the database. Internet access will remove this configuration restriction, as well as provide the data security needed so that only authorized users are provided access. This capability will allow a broader community access to the database.

3. Continuation of the current database system hardware/software configuration.

This area will continue to be pursued because it provides a high assurance of control of documents for security purposes. Also, the current setup maintains a capability while developing Internet capability at Arzamas-16 and Chelyabinsk-70. In addition, a backup capability of the database is desired in case any difficulties are encountered with Internet access.

4. Expansion of topics for inclusion in the database.

This area is important to all three institutes because it addresses the desire to attract future customers from a wider spectrum of safety-related issues, such as those in the oil and gas industry. For example, a potential Russian customer in the future may be Gazprom.

5. Investigation and implementation of translation language software

Currently, there is a wide variety of English/Russian translation software that needs to be evaluated in order to decide on a package that can easily be used by the greatest number of customers. The availability of a translation software package that is part of the database system and which can provide reasonable text translations into either Russian or English will enhance the appeal of the database to those customers who are not conversant in one or both languages. Hence, a practical goal of this cooperative endeavor is to have full text search capabilities in both English and Russian, as well as English-to-Russian/Russian-to-English software-generated translations.

CONCLUSION

While there is much work to be done in improving the US/Russia safety database, nonetheless much has been done by both Sandia and the two Russian institutes in developing the baseline data and the system hardware/software configuration. Considerable effort was expended on the part of the participants at all three institutes in order to overcome both the logistical and bureaucratic obstacles that stood in the way of creating this collaborative database. As a result, the very nature of this work has resulted in a greater spirit of trust and cooperation between American and Russian nuclear weapons engineers involved in this endeavor, and their willingness to share their information and experiences with each other will hopefully result in developing better safety approaches for low probability, high consequence engineered systems. In addition, the increased accessibility of information describing both Russian and American experiences in the safety area may be of potential benefit to the world community.

- Computer System
  - 486/66MHz - 8 MB RAM
  - 2-4 GB Hard Drive
  - 1MB Video Card (1024 x 768)
  - 230 MB Bernoulli Drive
- NEC 14-inch Monitor
- HP-4L Printer (300dpi)
- Copier
- Optical Drive Unit
- Scanner

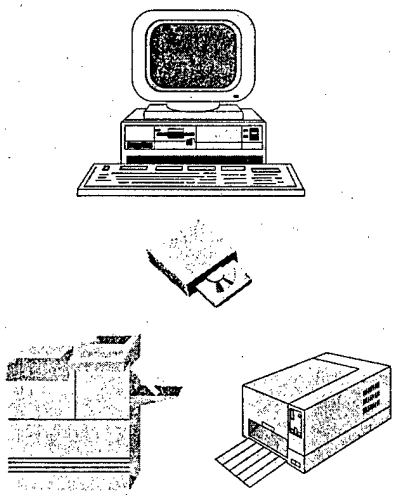

Figure 1. Computer system hardware used at Arzamas-16 and Chelyabinsk-70 in support of the US/Russia safety database.

ACKNOWLEDGEMENTS

The authors wish to thank Michael Krawczyk and Debra Buttry of Sandia National Laboratories for their continuing efforts in maintaining and continually updating the US/Russia safety database.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

REFERENCES

1. Microsoft, MS, MS-DOS, and Windows are registered trademarks of Microsoft Corporation, Redmond, Washington 98052-6399.
2. Ananiychuk, Vladimir N., Task 1 of Contract No. AK-5005 with Sandia National Laboratories: "Development of a Surety Information Center Database for High Consequence Engineered Systems," November 1995.
3. LaserFiche for Windows is a registered trademark of Compulink Management Center, Inc., 370 South Crenshaw Boulevard, Suite E-106, Torrance, California 90503.
4. LaserFiche for Windows User's Guide, Version 2.2, Compulink Management Center, Inc., 370 South Crenshaw Boulevard, Suite E-106, Torrance, California 90503.3
5. Stylus for Windows Version 2.21, copyright PROject MT, Ltd., 1993-1995, St. Petersburg, Russia 199053.
6. Volkova, Nina A., Final Report on Contract AK-5006: "The Development and Providing [for] Database on High Consequence Engineered Systems," April 4, 1997.
7. Volkova, Nina A., Report on the second stage of the AK-5006 Contract between Sandia National Laboratories and the Russian Federal Nuclear Center – VNIIEF: "Development of a Surety Information Center Database for High Consequence Engineered Systems," April 1996.
8. FineReader 2.0 Optical Character Recognition System, copyright 1993-1995, BIT Software, Inc., Moscow, Russia 105568.
9. Dvorack, Michael A., "Russian Risk Assessment Methods and Approaches," 14th International System Safety Conference, Albuquerque, New Mexico, USA, August 16, 1996.

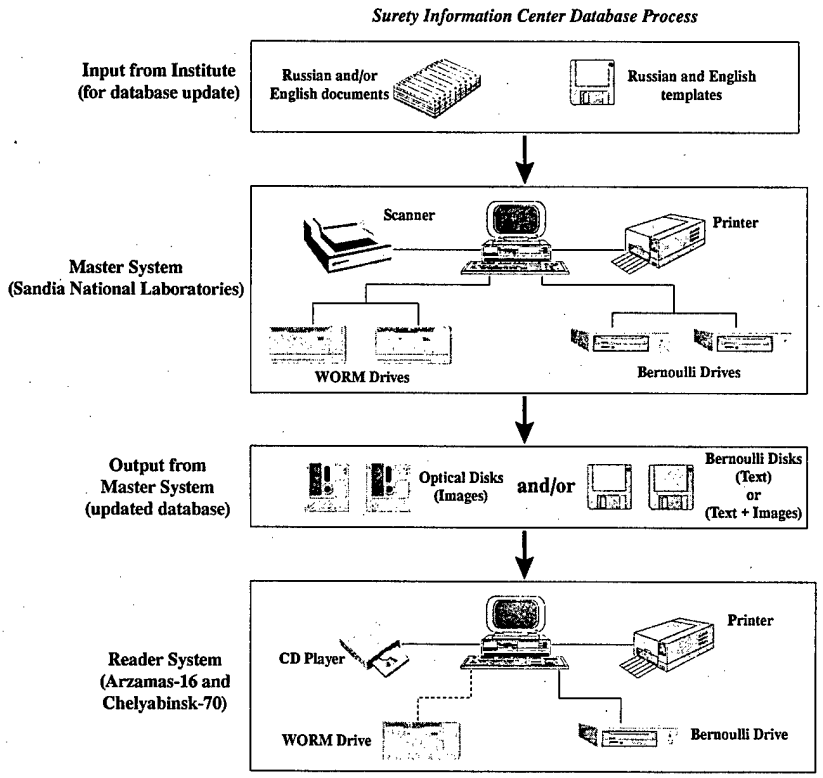

Figure 2. Process for maintaining and updating the US/Russia safety database.

Report Number (14) SAND--98-0247C
CONF-980621--

Publ. Date (11) 199801
Sponsor Code (18) DOE/DP , XF
UC Category (19) UC-700 , DOE/ER