

Building a Data Infrastructure for Conducting Research in Ukraine

Randolph Lucas, Pablo A. Destefanis,
Craig R. Hollingsworth, Mary-Anne E. Ardini-Poleske,
and Martha J. DeCain

September 2010

About the Authors

Randolph Lucas, MS, is a senior research programmer/analyst in RTI International's Research Computing Division. For the International Consortium for Applied Radiation Research (ICARR) project, he served as the data processing task leader overseeing the design, development, testing, and deployment of the infrastructure and applications.

Pablo A. Destefanis, MS, is a senior systems engineer for RTI's Center for Information, Communications, and Technology. He provides technical coordination and expertise to international projects in information systems.

Craig R. Hollingsworth, MA, is an IT/QA specialist in the Software Quality Assurance Group of RTI International's Research Computing Division. He specializes in system security documentation and National Institute of Standards and Technology (NIST) guidelines and controls as they relate to the system Certification and Accreditation (C&A) Process.

Mary-Anne E. Ardini-Poleske, BA, is a senior clinical studies director in RTI's Statistics and Epidemiology Unit. She was the ICARR Project Director.

Martha J. DeCain, BSBA, is a research health analyst in RTI's International Studies Program in Statistics and Epidemiology. For the ICARR project, she served as deputy director, overseeing the fiscal status of the project and subcontracts, and as business manager.

RTI Press publication RR-0012-1009

This PDF document was made available from www.rti.org as a public service of RTI International. More information about RTI Press can be found at <http://www.rti.org/rtipress>.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Lucas, R., Destefanis, P. A., Hollingsworth, C. R., Ardini-Poleske, M. E., and DeCain, M. J. (2010). Building a Data Infrastructure for Conducting Research in Ukraine. RTI Press publication No. RR-0012-1009. Research Triangle Park, NC: RTI International. Retrieved [date] from <http://www.rti.org/rtipress>.

This publication is part of the RTI Press Research Report series.

RTI International
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
Fax: +1.919.541.5985
E-mail: rtipress@rti.org
Web site: www.rti.org

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the US Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the US Government or any agency thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the US Government or any agency thereof.

©2010 Research Triangle Institute. RTI International is a trade name of Research Triangle Institute.

All rights reserved. Please note that this document is copyrighted and credit must be provided to the authors and source of the document when you quote from it. You must not sell the document or make a profit from reproducing it.

[doi:10.3768/rtipress.2010.rr.0012.1009](https://doi.org/10.3768/rtipress.2010.rr.0012.1009)

www.rti.org/rtipress

Building a Data Infrastructure for Conducting Research in Ukraine

Randolph Lucas, Pablo A. Destefanis,
Craig R. Hollingsworth, Mary-Anne E. Ardini-Poleske,
and Martha J. DeCain

Abstract

This report describes the IT infrastructure and health information exchange system (HIES) software developed for the International Consortium for Applied Radiation Research (ICARR) project. The research was conducted under a cooperative agreement funded by the US Department of Energy in collaboration with RTI International, Duke University, the University of North Carolina at Asheville, and the Research Centre for Radiation Medicine in Ukraine. The ICARR research consortium was formed to study how the human body responds to primarily low-level nuclear radiation exposure. A major component of the project was the creation of a cohort of workers hired to secure the deteriorating object shelter that surrounds Chernobyl Nuclear Power Plant reactor #4. These consenting workers provided medical data and biological samples to the consortium for research testing. A significant challenge to the project was setting up an information technology system for data capture, data transmission, and information sharing among the international consortium partners. This paper describes how project goals and the research environment shaped the IT solution that was successfully deployed within nine months of contract award.

Contents

Introduction	2
Building an IT Infrastructure	2
The Challenge	2
General Approach	3
Open Source Software	3
Vendor Selection and Procurement	4
Internal Network	5
Internet Connection	6
Network Security	6
Server Configuration	6
Backup, Monitoring, and Training	7
Building the Data System	8
The Challenge	8
Planning and Design	8
Features of the HIES	8
Conclusions	12
Acknowledgments	Inside back cover

Introduction

In late 2006, under Congressional mandate, the US Department of Energy issued funding to a partnership led by RTI International to study how the human body responds to primarily low-level nuclear radiation exposure. The partnership, composed of RTI, the Research Centre for Radiation Medicine (RCRM) in Ukraine, and Duke University and the University of North Carolina at Asheville (UNC-A) in the United States, established the International Consortium for Applied Radiation Research (ICARR), under whose auspices research projects on radiation health impacts would be conducted. The first project undertaken was the Chernobyl Research and Services Project (CRSP), a data collection effort conducted in Ukraine. RTI was tasked with designing, procuring equipment and software for, and installing a working IT infrastructure and data capture system at RCRM, a government facility in the Ukraine. The data system was called the health information exchange system (HIES).

Our partner, RCRM, a member of the Ukrainian Academy of Medical Sciences, was established in October 1986 in response to the Chernobyl nuclear accident, which occurred on April 26 of that year. RCRM has treated victims of the accident, including the emergency workers who responded to the original disaster. They have conducted research on the effects of the accident and now conduct medical surveillance and worker safety screening for the reconstruction effort taking place at Chernobyl Nuclear Power Plant (ChNPP). Through HIES, we collected information on the health of the workers rebuilding the protective structure that contained the remains of the exploded reactor. The project also collected blood samples from participants; these samples were used by the project consortium to perform genetic analysis to identify gene expression changes that may be associated with radiation exposure. Data related to the samples and testing was also a part of HIES.

Our team had a short timeline and a limited budget, and most notably had to do most of its design and development long distance (6,000 miles away) without adequately trained staff in Ukraine to assist. The unfamiliar cultural environment, deteriorated

Soviet-era buildings at RCRM, and limited technology infrastructure in Ukraine added to the complexity of the assignment.

The RTI HIES team, from the start, was aware of the challenges in establishing a system to specification within the project constraints. The design called for a new physical information technology (IT) infrastructure in Ukraine, including a local area network (LAN) with servers; a mirror server in the United States at our campus in Research Triangle Park, North Carolina, for project monitoring and data processing and storage; and a data collection interface that could be used on both servers. We had the challenge of ensuring that both servers contained real-time data, with transfers scheduled daily, for both backup and research purposes.

Building an IT Infrastructure

The Challenge

Developing an IT infrastructure is always challenging. Even when installing a small network in your home, you face constraints with the design and construction materials of the house, the chosen equipment, needs of other people in your household, and your budget.

The RTI team grappled with these same issues on a huge scale when building the network in Ukraine. We also faced additional challenges:

- The need to work from the United States, seven time zones away, to investigate available resources
- The need to involve inexperienced local IT staff in Ukraine
- The need to present recommendations on purchases in a manner that permitted the principal Ukrainian partners to be part of the decision-making process
- The need to understand and respect local business customs (i.e., “no return” policy) and language impacts.

The team also had to weigh factors like equipment life cycle, the need for redundancy and integration of existing older equipment, and ongoing maintenance and support. We had to adapt to the physical layout and structure of the RCRM buildings, which were constructed with thick concrete walls during the

Cold War to survive weapons attacks. The walls contained no space for cabling, having been designed before such needs existed. None of these individual challenges was insurmountable, but taken together, they added difficulty to a research project that had little start-up time built into its schedule.

General Approach

We determined immediately that we would purchase equipment and services in Ukraine to support local business and, through these contacts, build a sustainable and expandable infrastructure. We designed the network configuration and system to address to the needs of the project (i.e., creating a specimen database) but also to satisfy broader requirements to build capacity at RCRM (i.e., internal e-mail server). We used a “best value” approach to avoid unnecessary equipment purchases and used a conservative rationale for configuration options to provide equipment that would, without major upgrades, have a minimum life-cycle of five years. We provided redundancy for core components such as servers, and assigned secondary roles to the existing standby equipment to make full use of the available IT resources. We also decided to use open source technology for a number of key components.

Open Source Software

We decided to use open source software for the e-mail, web server, and data system itself for a number of reasons. Overall, the team found no negatives to using open source software. We saw international regulations restricting and controlling transfer of software as potential problems for the project in the future. Open source allowed for future low-cost expansion. We also knew that open source solutions are readily supported by many developers internationally.

We identified several drawbacks to using a commercial, proprietary software package. Commercial software products can require a support contract that includes training and minimum feature modification support. Enhancements or fixes may be released only on a schedule set by the vendor. Because they are often not customizable, they can be less agile and require time to reach the level of development required by a project. Also, software firms are driven

by market concerns—creating upgrades that raise costs and using proprietary formats that minimize export compatibility.

Open source solutions are not tied to a particular vendor, which gave the project options for changing the support structure over time if necessary. The lack of ownership by a specific manufacturer means that open source solutions are not driven by marketing pressures that would affect the project budget both immediately and in the future. Finally, we felt that open source solutions are geared toward flexibility and interoperability. If at some point the project decided to move to a proprietary solution, the open source standards would facilitate this move more readily than the reverse.

In particular, the web server and e-mail server were developed using open source technology to limit cost and help with license management. In both instances, the software cost was kept low because there were no licensing fees, which in turn meant no maintenance fees for the project.

We obtained the software and installed it as many times and in as many locations as we needed. We had no need to count, track, or monitor copies for license compliance, which simplified management for project personnel. Furthermore, we faced no restrictions on transferring the license to our partner organization at the end of the project.

Regarding the HIES data capture application itself, we weighed the cost of developing a new application against the cost of buying and customizing a packaged data-capture program. A commercial package would have had the benefit of many hours of testing behind it. However, the application built on that toolset would have required the same level of testing to validate it for production use as those applications built using open source software. Most software vendors provide support contract services. By using open source technology, we recognized that the project would need to engage a local developer to continue to support the application after RTI’s involvement ended. However, as noted above, we felt that open source developers were plentiful and available to respond.

We should note that not all the software we used was open source. To inventory the specimens collected from workers, we used a highly specialized proprietary application called Freezerworks that was designed to work with specific hardware. For example, the application interfaced with a bar code printer that printed labels specifically designed for use on small vials stored in a -80°F freezer. Development of this very specialized functionality in the open source toolset was not a reasonable option considering the cost and project schedule. We also used commercial software (Windows XP) for the desktop operating system because it was familiar to the majority of users.

Vendor Selection and Procurement

Long before we laid cables and installed servers, we faced less tangible challenges. Our first major decision was the choice of hardware vendors from whom to procure the computer equipment and who could provide the additional technical support needed to build the IT infrastructure. Working with RCRM, we selected vendors in Kyiv to provide and install equipment and to help with the initial configuration of the components.

We hoped that by working with local vendors, we would establish a relationship to provide support to sustain and expand the infrastructure on a long-term basis. We also knew that the vast language and time zone differences could make working with vendors outside of Ukraine more complex, time-consuming, and expensive. Finally, we discovered that the length of time to procure equipment outside of Ukraine and successfully clear customs would likely prevent us from hitting our target launch date.

E-mail was the most productive means of communicating throughout the procurement process. Kyiv is seven hours ahead of Eastern Standard Time, where the RTI home office is located. RTI staff could pose questions, and our partners in Ukraine could find the answer and send a response, which would arrive by the time the workday began again in North Carolina. However, early on it became apparent that we needed questions translated as we communicated with vendors. We found that even people who

understood English misunderstood the nuance of some of our questions. Having our communications translated produced delays, but fortunately RTI had a project office in Kyiv with bilingual staff well-versed in the project's needs. These staff facilitated when needed, translating and making phone calls to clarify information.

We retained two hardware vendors, both of whom had worked with RCRM in the past, to assist with the work. Each represented a different company style: the first was a smaller, local company; the second a large, international company. The smaller vendor wanted only to provide hardware, and lacked the interest or capacity to provide consulting services or a comprehensive solution. The larger vendor attempted to provide an integral solution, covering the network design, equipment, and installation. While the ability to provide overall services was not the main evaluation criterion, it carried weight in our ultimate selection.

The three aspects we considered for each company were experience, cost of hardware, and timeline for installation. The small, local vendor offered a lower price, and RCRM had had a previous positive experience with them, so they were our first choice. However, we tested the networking equipment available through them and found repeated faults, and the vendor had a slow response time in correcting them. We therefore chose the larger vendor to perform the network cabling and provide the telecommunications equipment.

However, working with the larger provider was not trouble-free. The vendor strongly advocated that we purchase equipment that was not consistent with our "best value" approach. The issue was further complicated by accepted professional protocols in Ukraine, which dictated that consultations with the vendor had to be handled by RCRM leadership staff, who did not have a background in information technology. These indirect discussions between the vendor, RCRM management, and project IT staff about the equipment selection criteria and the network design delayed the installation and almost caused us to miss the target implementation date.

We found, however, that our decision early in the process to choose the larger vendor for the installation was correct and saved many problems. We bought our servers from the smaller vendor and the communication equipment from the larger vendor, and had no major problems with either. Because the same vendor provided network equipment, cabling, and installation services, the server work was completed with no dilution of responsibilities. Their personnel laid out the approach well, and they installed cabling neatly and rapidly; their work interfered minimally with the day-to-day operations of the facility. We were also presented with complete test results and maps of the cabling work, which reflected the vendor's high quality of work and would help the project in the future.

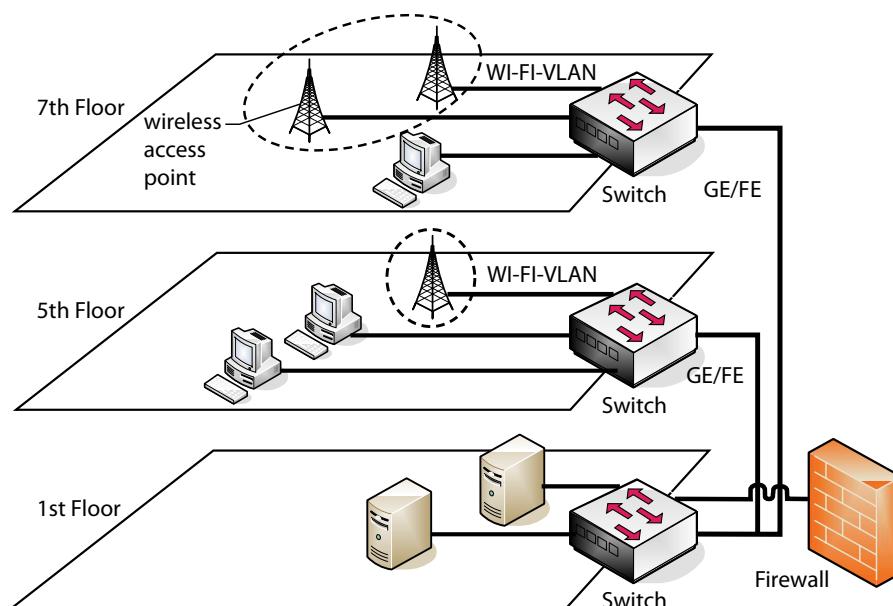
Internal Network

The existing network infrastructure in Ukraine was composed primarily of the remains of ad hoc installations carried out by several previous unrelated research projects. The approach to networking had not been centralized, and the equipment was consumer-grade, scattered in various locations throughout the building, often not interconnected. The cabling followed the same approach, with overlapped installations and no centralized patching.

To not repeat this shortsighted approach, we decided to implement a network infrastructure that enhanced the capacity of RCRM as a research facility rather than addressing the bare minimum requirements for our project. RCRM was housed in government buildings from the Soviet era, which had not been upgraded in decades. In addition, RCRM had no funding in its operational budget to invest in computer technology, leaving well-trained scientists without tools that would help advance their work. One of the strategic goals of ICARR as a partnership was to address these limitations.

To increase IT capacity, we designed a collapsed backbone network, with one main network switch in a refurbished data room and secondary network switches on other floors. To stay within budget, we provided only floors on which the study would take place with a network switch (Figure 1). In the future, however, RCRM IT personnel can install a network switch and run cable to where it is needed. Once this is done, any newly installed devices can connect with the RCRM networked devices. This flexibility extended the reach of the network and allowed connected users to share existing resources, such as an Internet connection.

Figure 1. Research Centre for Radiation Medicine (RCRM) network layout



Wi-Fi: A wireless connection using one or more of the 802.11 standards

VLAN: Virtual Local Area Network: A logical division of a local network used to isolate groups of computers

GE: Gigabit Ethernet: An interface with maximum capacity of 1 billion bits per second of throughput (1 Gbps)

FE: Fast Ethernet: An interface with maximum capacity of 100 million bits per second of throughput (100 Mbps)

RTI worked closely with the RCRM building engineer to plan the cabling work. The building engineer provided layouts and advice in routing cables. This assistance was essential because the building where most project tasks took place was a concrete structure with thick brick and concrete walls with no space for data cabling. Drilling through these walls was a very difficult, time-consuming task that took place in areas where medical examinations were conducted. The installation crew had limited time on any given day, and thorough clean-up was vital. We routed all the cables externally, protecting them with plastic ducts, except the links between floors. This approach adequately balanced cost, simplicity of installation, and aesthetics.

The building construction characteristics restricted the reach of the wireless units. In most cases, the units could barely be accessed from outside the rooms where they were located. However, the units were meant to connect mobile users at designated spots, rather than create a fully wireless environment. This has not become a problem for RCRM.

In total, 50 cable runs were added on three floors, covering all the areas used by the project and the offices of senior staff members. We provided a minimum of two connections each, to allow for future expansion. This approach provided RCRM with a functional, reliable, and scalable IT infrastructure. In addition to the wired infrastructure, CRSP provided RCRM with four wireless hotspots, strategically located in areas where senior staff meetings are held, to provide for uninterrupted connectivity. We made this IT foundation available to other projects as well as the medical units, and RCRM can incrementally improve the existing structure deliberately and cohesively.

Internet Connection

We signed an initial contract with a local Internet service provider to provide a wireless radio frequency link from the project servers to the Internet. At the time of this decision, the viable solutions were a wireless link or a dial-up link; DSL services became a possibility only later in the project. While the wireless link was stable (availability >90 percent), it proved

sensitive to weather conditions. Also, the cost was based on total Internet traffic used, and RCRM faced high usage bills as more users became aware of this service and connected to the Internet.

For a more reliable connection, and to limit communication expenditures for the project, we switched to a DSL-based service when this became available to RCRM. The service proved easier to maintain, had a higher availability, and had a flat-rate plan. The contract was signed for a 512 Kbps line, with the option of adding or subtracting capacity as required by the project and RCRM priorities.

Network Security

The RTI team performed a risk assessment on the system and separated access to RCRM information assets into external and internal threats. External threats were defined as any traffic arriving through Internet connections, and internal threats were defined as coming from devices connected to the RCRM network. We managed external threats with two firewalls. The first was a Cisco ASA5500 unit sitting between the Internet router and the internal network. The second was based on iptables software and was set up on each project server. Both firewalls needed to verify the incoming connections to a server-based resource before allowing access. To simplify the operation of two firewalls from different vendors, we managed both units with a single interface generated by a graphical tool.

Server Configuration

The project team used a vacant space in the lower level of RCRM as a data room. The room was renovated with electrical connections and air conditioning. CRSP personnel installed and configured two servers each in an enclosed rack with independent uninterruptible power supplies (UPSs). While only one server was required to host the project data system applications, databases, and ancillary services, we provided redundant servers to ensure continuous connectivity between Ukraine and the United States, in case one unit failed.

The main server supported both the web server and the main database, with a complete copy of the data

held on the secondary server. If the main server failed for any reason, the second server would perform the primary functions. We ensured the data replication was automatic and required no manual intervention under normal conditions. RTI personnel also monitored the process in the United States.

We set up a virtual server installation on the secondary server to allow both servers to be used to their potential under normal operation. This allowed RCRM IT staff to host virtual servers to test new systems and to practice general IT processes without interfering with the production servers. This arrangement gave the RCRM staff an opportunity for computer training; most training had been curtailed by the limited funding provided to governmental institutions in Ukraine.

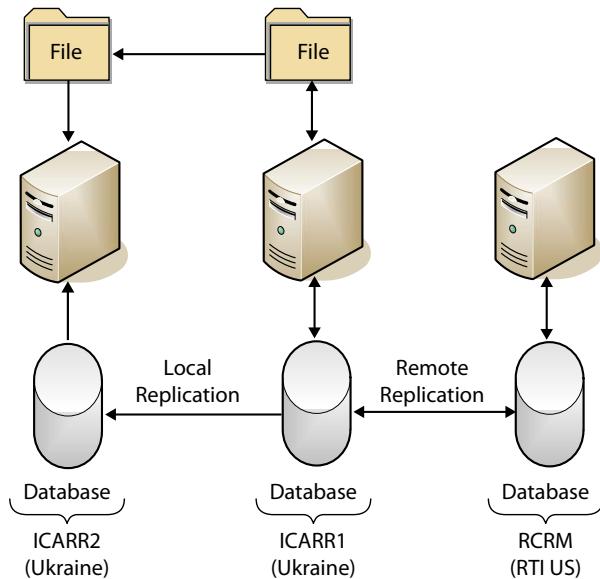
The servers housed a database and web server for the data system application as their main task. We were able to also add other services consistent with our philosophy of providing basic IT capabilities in a centralized manner. For example, we added a mail server, which allowed RCRM to have a single set of accounts and reduce the use of outside e-mail providers. Outside e-mail providers add to security risks because they are potentially less secure, and they consume more bandwidth than solutions hosted in-house. We also accommodated traveling users by developing and installing a web-based interface.

Backup, Monitoring, and Training

After we instituted the server-to-server database replication (Figure 2), we ensured that the data were backed up to a local hard drive and transferred nightly to an RTI server in North Carolina. This data transfer acted as a secure, off-site storage method, a feature we had not instituted at RCRM.

RTI provided basic server monitoring services for the project; IT staff could quickly view the status of the database operations and the database replication between the servers in Ukraine and the US. We designed the setup to be nonintrusive and not require

Figure 2. Server-to-server data replication



intervention by RCRM staff. Our ability to monitor the replication process from a distance provided the project with some early-warning capabilities. We used a monitoring tool that is open source and freely available.

Because funding for IT building and staff training at RCRM was extremely limited, we involved RCRM IT staff in the capacity building process from the start. Our IT personnel worked to familiarize the RCRM IT staff with each component and its role in the project. We installed and configured the network together with the RCRM staff, using a translator to help. We also replicated the project data system to the secondary server to give RCRM IT staff a virtual environment for practice purposes (what is commonly referred to as a “sandbox configuration,” a mirror image of the production environment, where the administrators can practice and experiment without disrupting normal operations. In addition to this on-the-job training during the life of the project, RTI hired local training centers to provide short courses (five to ten days) on the project’s operating system and database engine.

Building the Data System

The Challenge

While IT personnel were planning and building the infrastructure, project programmers were designing in parallel a software program to capture the data that would be the foundation of the ICARR research resource. Our overriding challenge from a data development standpoint was the project timeline: implementation at eight months from award. This compressed timeline required us to develop the IT infrastructure and the data system at the same time, processes that are normally staggered.

Compounding the challenge, debate waged over which items were best for data collection while design of the system for data capture was pressed to move forward. We therefore had to continuously revise and remained uncertain about the ultimate look and feel of the system from a user perspective. Despite the constant development and revisions, we deployed HIES in May 2007; it became operational on June 1, 2007, and began collecting data immediately.

Planning and Design

In mid-October 2006, team members from RTI and Duke University traveled to Kyiv for a kickoff meeting with our RCRM partners. We spent many hours inspecting the clinical facilities, guided by RCRM personnel, to understand the research environment and to gather requirements. Our tours included the physical infrastructure of the buildings themselves, especially the features that were going to affect how we installed the equipment and instituted the LAN. For the system platform, we considered initial cost, long-term licensing and support costs, and international software transfer regulations. For our partners in Kyiv, long-term costs and licensing were key issues.

We could have purchased over-the-counter software in Kyiv, but the purchase would have required another requisition process similar to that needed for hardware and network equipment procurement. The RTI team felt this would have jeopardized the schedule and budget. Both RTI and RCRM personnel

agreed that open source technology could address the immediate needs of the project while providing a low-cost method for RCRM to expand its capacity. We aspired to build a system based on standard computing practices in health research. An open source platform provided a means by which we could apply those standards to the systems development in an efficient, cost-effect manner, including instituting 100 percent key-verified input, which helped to control the quality of the data collected. Using an open source platform proved very successful: the system was developed and installed at the RCRM on schedule without going over budget, or invoking regulatory issues. For the database architecture, which had to be robust, the RCRM team wanted to use Oracle as the database engine. However, after discussion and consideration of cost issues, we agreed to use MySQL, which proved suitable for the data load.

Features of the HIES

We developed the HIES web portal on a Linux-Apache-MySQL-PHP platform at two server locations, in Kyiv and the US. The project required servers in both location due to the split in responsibilities (data collection in Ukraine and data monitoring in US). We developed the portal to be a flexible and powerful tool for data entry, reporting, and extraction and as a conduit for project information across institutions. We ensured that the web portal at each facility ran the same software, and staff access was controlled by user role and username/password. We also synchronized the databases so the US project team always had access to the most current clinical data.

Dual Language System

To accommodate the English and Ukrainian speaking team members, we provided a language toggle button to switch the HIES web portal menus between English and Ukrainian (Figure 3). We worked closely with RCRM staff as they translated the English portal information into Ukrainian.

Web-Based Data Capture

Both RTI and RCRM agreed that a key component of HIES was data capture via a web-based data entry interface (Figure 4). We designed the tool to apply range and consistency checks that produced high-quality data needing limited cleaning before analysis.

Laboratory Sample Tracking

HIES imported data from Freezerworks, the commercial specimen inventory system the project used to manage the sample collection and processing component. Importing the data from Freezerworks allowed HIES to check the clinical and laboratory data for consistency and completeness (Figure 5).

Figure 3. HIES welcome screen with bilingual support



Figure 4. HIES sample data entry screen

 A screenshot of the 'Health Information Exchange System' data entry interface. The top bar shows the title and the user 'Brenda B. Lusk' with the role 'Administrator'. The main window has a table with columns for 'Question' and 'Response'. The 'Question' column contains questions such as 'ICARR Project', 'Дата обстеження' (Date of Exam), 'Код робітника / Subject ID (ICARR)', 'Is there a Pulmonary diagnosis?', 'First Pulmonology diagnosis', and '3.1 МІОК-10 код' (3.1 ICD-10 code). The 'Response' column contains input fields and dropdown menus. The left sidebar shows a navigation menu with items like Home, Announcement, Wiki, Documents, Schedule, Calendar, Reports, Admin, Projects, MyProfile, Login/Logout, Feedback, Tools, Uri, ToolAdmin, Test Tool, AliquotType, Pul_Diagnosis, Blog, Project Contact, and GuestBook. At the bottom, there are buttons for Save, Report, Help, Bug, Print, Back, Delete, and a status message 'Meds: 0 Status: Can't find'.

Figure 5. HIES check screen

Tab:	Biospecimen Sample Information														
Title:	Інформація про біозразки														
Detail:	As of: 2008-10-02														
	Input Control														
	RTI														
	<table border="1"> <thead> <tr> <th>AliquotType</th> <th>Number</th> </tr> </thead> <tbody> <tr> <td>Buffy Coat (EDTA)</td> <td>454</td> </tr> <tr> <td>Plasma (EDTA)</td> <td>1356</td> </tr> <tr> <td>Residual RNA</td> <td>113</td> </tr> <tr> <td>Wholeblood (EDTA)</td> <td>227</td> </tr> <tr> <td>Wholeblood PAXgene</td> <td>77</td> </tr> <tr> <td>Wholeblood Serum</td> <td>1362</td> </tr> </tbody> </table>	AliquotType	Number	Buffy Coat (EDTA)	454	Plasma (EDTA)	1356	Residual RNA	113	Wholeblood (EDTA)	227	Wholeblood PAXgene	77	Wholeblood Serum	1362
AliquotType	Number														
Buffy Coat (EDTA)	454														
Plasma (EDTA)	1356														
Residual RNA	113														
Wholeblood (EDTA)	227														
Wholeblood PAXgene	77														
Wholeblood Serum	1362														
	RCRM														
	<table border="1"> <thead> <tr> <th>AliquotType</th> <th>Number</th> </tr> </thead> <tbody> <tr> <td>Buffy Coat (EDTA)</td> <td>84</td> </tr> <tr> <td>Plasma (EDTA)</td> <td>2170</td> </tr> <tr> <td>Wholeblood (EDTA)</td> <td>371</td> </tr> <tr> <td>Wholeblood PAXgene</td> <td>370</td> </tr> <tr> <td>Wholeblood Serum</td> <td>1939</td> </tr> </tbody> </table>	AliquotType	Number	Buffy Coat (EDTA)	84	Plasma (EDTA)	2170	Wholeblood (EDTA)	371	Wholeblood PAXgene	370	Wholeblood Serum	1939		
AliquotType	Number														
Buffy Coat (EDTA)	84														
Plasma (EDTA)	2170														
Wholeblood (EDTA)	371														
Wholeblood PAXgene	370														
Wholeblood Serum	1939														

Reporting

HIES provided a set of dynamic reports in English and Ukrainian on worker accrual, data entry status, fail-edit, and other process-oriented information. Additionally, we developed some basic clinical summary reports to assist in quality assurance (Figure 6). Staff accessed response set distribution reports with a focus on ICD10 reported values from the web portal interface.

Data Editing and Error Resolution

Authorized personnel with the role of data manager could use the error resolution tool to edit incomplete or inconsistent data (Figure 7). This tool logged all changes to the data, and data managers could post comments explaining why they had changed the value. Most often, the managers changed data after reviewing the original paper forms.

Data Access

Users could also select tables, specify filters, and set sort orders to generate lists from the data repository using the HIES database query tool (Figure 8). The system generated the lists as HTML, which the user could save to disk or print. Project staff also used this tool to export the data they selected to CSV (comma-separated values) format to import into other data analysis systems.

Figure 6. HIES reports screen

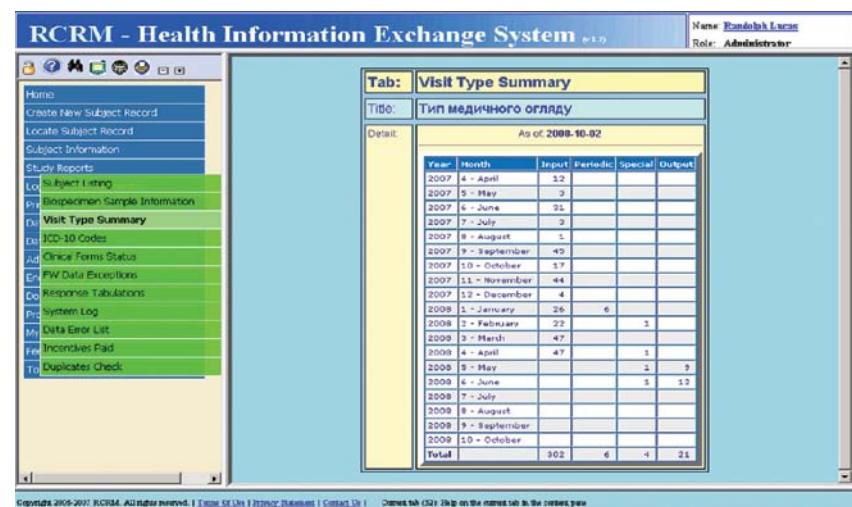
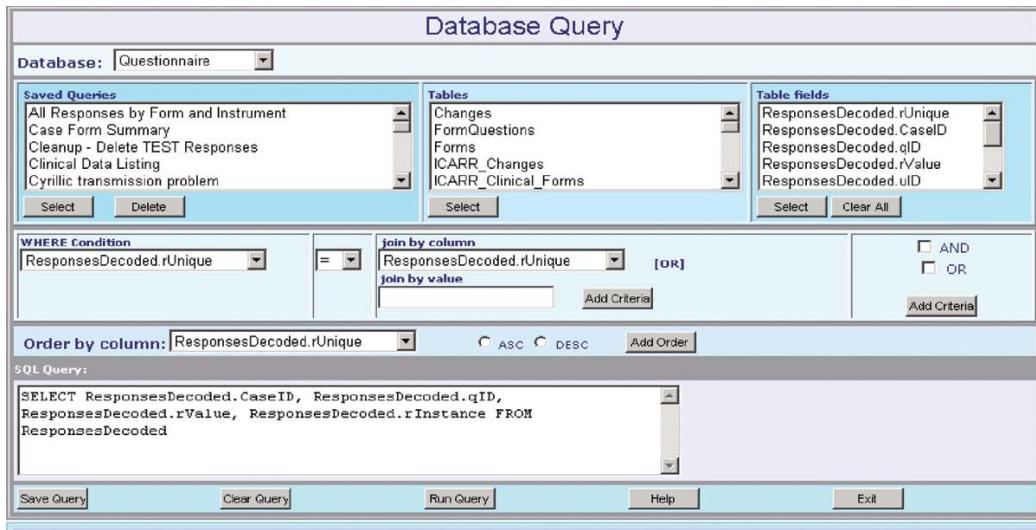


Figure 7. HIES edit resolution screen

Edit Resolution		Instructions
rUnique	15231254	The Edit Resolution form allows you to work the selected case for which an edit failure has occurred.
CaseId	10001130407	You should:
fID	Pul_History	<ol style="list-style-type: none"> 1. enter the appropriate value for the problematic response value and a 2. reason for the change and then 3. click on UPDATE
qID	pul2_cigpdday	
Question	5.0 Кількість випалених цигарок на добу Number of cigarettes smoked per day	
Visit Type	вхідний / input	
Visit Date	2007-04-13	
Old Value	8 - 20	
New Value	<input type="radio"/> до/less than 8 <input checked="" type="radio"/> 8 - 20 <input type="radio"/> більше/over 20 <input type="radio"/> Not Available	
Rule	if (GetVal('pul2_smkstatus') != '1' and GetVal('pul2_smkstatus') != '') \$stat=Required('\$val'), elseif (GetVal('pul2_smkstatus') == '' or GetVal('pul2_smkstatus') == '1') \$stat=IsBlank('\$val'), else \$stat=1,	
Reason	<input type="text"/>	
Update		The new response value will be inserted into the Responses table

Figure 8. HIES database query tool screen



We included a set of web-based tools that allowed authorized staff to manage users, roles, and content (Figure 9). These tools also enabled RTI to support the RCRM system in real time; we were able to set up new users, introduce new tools, fix bugs, view data, and inspect system logs.

Bug Tracking and Feedback

HIES offered a feedback tool to note bugs, request help, and suggest new features (Figure 10). The tool notified systems support immediately via e-mail and sent a confirmation e-mail to the user logging the issue. We integrated this tool with the data capture tool to help identify specific scenarios where users encountered problems during data entry. When the user sent a bug report from the data entry interface, HIES automatically noted the subject ID and form report, so the user only had to add his or her comment. After the system administrator received the automatic e-mail, he or she was able to address the bug

Figure 9. HIES configuration management screen

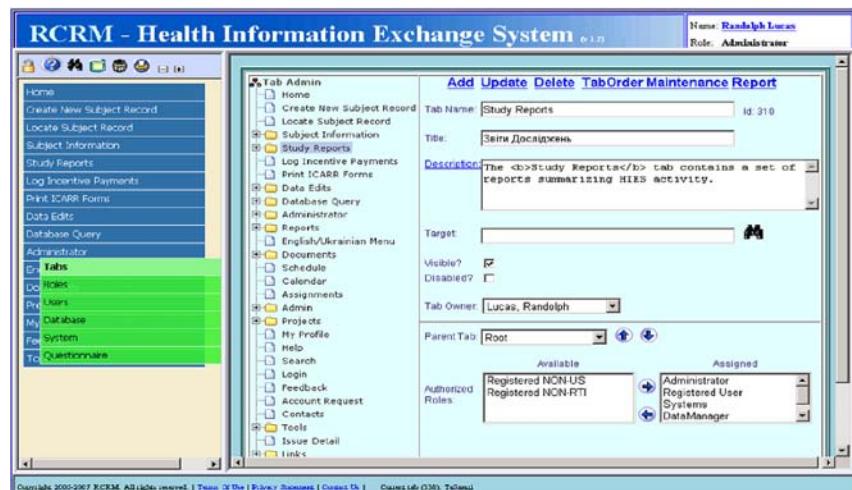
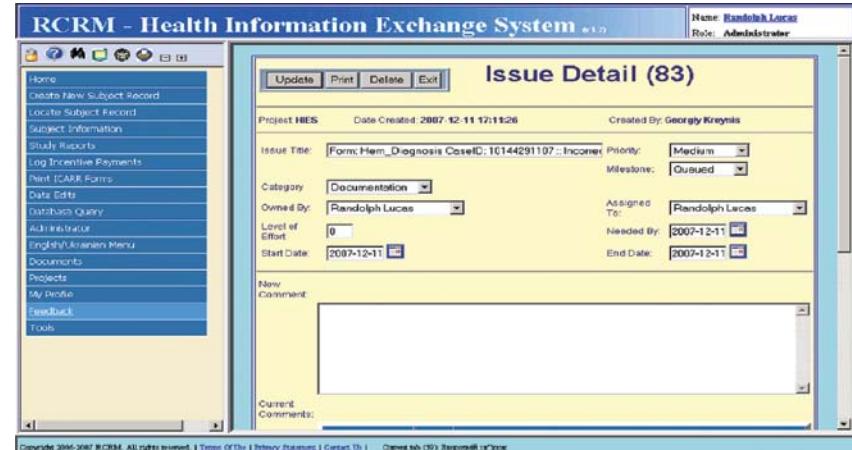


Figure 10. HIES issue reporting screen



within 24 hours. When the problem was resolved, the administrator sent an e-mail to the user who reported the problem. Translation by RTI staff was available so that users could submit bug reports in Ukrainian or Russian as well as English.

Data Exchange

The HIES had dual databases at RTI and RCRM. A data exchange process synchronized the two and transferred data each night; however, unresolved edits and unverified data were not transferred. The interface between the RTI server and the RCRM server was built on scheduled PHP scripts that extracted and exported the data from Ukraine and imported it into the US database. At first, the replication process copied the entire database from one server to the other, which took several hours due to the slow speed of the Internet and erratic connectivity in Ukraine. The copying process interrupted normal data entry at RCRM because it used such a large amount of system resources. Also, because the system was busy performing the database copy, no automated alerts were sent that identified the problem that was interrupting data entry. We soon revised the replication process to copy only changes from one database to the other, a much faster process that produced no problems.

Conclusions

The project team worked to anticipate problems we knew would occur, such as power failures, Internet connectivity downtime, equipment failure, and others. For each problem, we tried to implement a solution in advance: We installed an uninterruptible power supply (UPS) to deal with regular power

outages. We addressed breakdowns in the Internet connection by designing our data transmission system to automatically synchronize the data between servers when the service returned. We installed redundant servers at the RCRM to minimize the downtime from server crashes. Ultimately, we knew that the system's dependability was linked to the local ISP service and to the reliability of the power grid infrastructure in Kyiv, both of which were sporadic. In particular, the harsh winter weather in Ukraine affected both systems. However, HIES functioned as intended and the project never lost data.

While creating the infrastructure for immediate use was a challenge for our team, we always made choices with the future in mind. An important part of the task was creation of a system that would remain functional beyond the contract period. This meant it had to be reliable, durable, and expandable as RCRM needs changed and IT capacities deepened. We met this challenge by using affordable open source programs, solid hardware, and a logical design. Knowing that documentation is an important component to ensuring the long-term use of a system, RTI developed an operations manual, user guide, and technical specifications documentation that detailed every aspect of the system.

Though we faced significant obstacles that impeded our progress and could have easily caused us to miss our deadline, we met our goal by working with our RCRM colleagues effectively within cultural, budgetary, and technological constraints to build an enduring IT infrastructure at RCRM that will continue to support RCRM's research on radiation exposure.

Acknowledgments

The authors would like to acknowledge the assistance and support of the US Department of Energy; the International Consortium for Applied Radiation Research (ICARR); the Research Centre for Radiation Medicine (RCRM); the Duke University Centre for Genomics; the University of North Carolina at Asheville; and RTI International employees Niamh Darcy, Susanna Cantor, and George Grubbs, coauthors of the HIES System Design document.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory and chemistry services.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International. More information about RTI Press can be found at www.rti.org/rtipress.