Ecotoxicology, 14, 275–281, 2005

Developing a *Cyberinfrastructure* for Integrated Assessments of Environmental Contaminants

TARANJIT KAUR,[1,5,*] JATINDER SINGH,[1] WING M. GOODALE,[2] DAVID KRAMAR[3] AND PETER NELSON[4]

[1] *Virginia Polytechnic Institute and State University Blacksburg, VA 24061, USA*
[2] *BioDiversity Research Institute, Gorham ME 04038, USA*
[3] *University of Maine, Orono, ME 04469, USA*
[4] *Advanced Computer Resources Corp, Nashua, NH 03063-1301, USA*
[5] *Department of Biomedical Sciences and Pathobiology, Virginia–Maryland Regional College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Phase 2 (0442) Duck Pond Drive, Blacksburg VA 24061, USA*

Accepted 4 December 2004

Abstract. The objective of this study was to design and implement prototype software for capturing field data and automating the process for reporting and analyzing the distribution of mercury. The four phase process used to design, develop, deploy and evaluate the prototype software is described. Two different development strategies were used: (1) design of a mobile data collection application intended to capture field data in a meaningful format and automate transfer into user databases, followed by (2) a re-engineering of the original software to develop an integrated database environment with improved methods for aggregating and sharing data. Results demonstrated that innovative use of commercially available hardware and software components can lead to the development of an end-to-end digital *cyberinfrastructure* that captures, records, stores, transmits, compiles and integrates multi-source data as it relates to mercury.

Keywords: cyberinfrastructure; mercury; database management; Portable Data Assistants (PDAs); field data collection

Introduction

Scientists conducting field studies on both biotic and abiotic indicators of environmental mercury contamination tend to rely on hand-written field notes and have only used digital technology for data storage and analysis (Smith, personal communication). Manual data entry into flat files or relational databases has resulted in relatively slow collation of paper-based records (Smith, personal communication). Although similar biological databases may exist or be developed by researchers with a common purpose, each database effort typically functions as a separate unconnected entity that has not been designed to be linked or cross-referenced to make multi-source data retrievable for aggregation and sharing (Stein, 2003). Furthermore, manual or digital efforts to integrate multi-source databases are generally hindered by a number of different factors, such as inconsistent data recording standards, lack of standardized nomenclature, inconsistent data collection formatting and differing database

*To whom correspondence should be addressed:
Tel.: + 540-231-6522; Fax: + 540-231-6033;
E-mail: taranjit@vt.edu

structures. The difficulties of integrating databases were evident during the process of assimilating the data for these issues of Ecotoxicology (Evers and Clair, 2005). While specific challenges varied from paper to paper, such as differences in units of measurement, geographic coordinates, and sample preparation, most authors spent significant time standardizing datasets. If, during their studies, researchers had access to a *cyberinfrastructure*, the process of sharing and coordinating results would have been significantly more efficient.

The term *cyber* suggests a very broad information point of view that can be achieved using information technology (Bajcsy, 2001). The term *cyberinfrastructure* refers to the digital computation, communication, information facilities and services needed to support the scientific and engineering research communities (Bajcsy, 2001). It is heavily based on computers, communications and storage technology, which distinguishes it from the traditional physical infrastructure. Thus, the development and use of *cyberinfrastructures* could revolutionize the way scientific research communities record, process and analyze data.

Mercury contamination has been identified as a significant environmental threat globally and at local, regional and national levels, particularly in northeastern North America (NESCAUM, 1998; Wiener et al., 2003; Evers and Clair, 2005). Effective management of environmental contamination requires the collection, rapid processing and interpretation of vast quantities of data relating to emissions from point and non-point sources, as well as an understanding of the underlying distribution dynamics (Kamman et al., 2004). Collaborative studies are important for effective and efficient management, intervention and policymaking as diverse environmental, meteorological, hydrological and geographic parameters, as well as the biotic and abiotic components of an ecosystem are fundamental to data analyses and integrated assessments (Evers et al., 1998; Wolfe et al., 1998). Multiple initiatives have been organized and are underway by government organizations to reduce environmental mercury emissions in the United States during the next decade (Smith, personal communications). However, there remains a need for better tools and methodologies for tracking and coordinating complex data on key environmental indicators, such as the common loon (*Gavia immer*), in order to better monitor and forecast the impact and effectiveness of these initiatives.

The objective of this study was to develop a *cyberinfrastructure* that would effectively and efficiently capture, record, transmit, store, compile and integrate multi-source data on key environmental indicators of environmental pollutants, specifically mercury. This paper describes the process used to design, evaluate and implement prototype software for integrating database management and geographic information for the purposes of aggregating and sharing data, as it relates to mercury, for integrated assessments by collaborating organizations.

Methods

This study was performed in four phases between 2001 and 2004. It was conducted by academicians and computer scientists, partnering with environmental biologists and mercury experts in governmental and private non-profit organizations. The field study test site was the Rangeley Lakes region of western Maine.

Phase 1. Program development

During this phase program objectives and needs were delineated. Several active researchers from academia, governmental organizations and private non-profit organizations were contacted and requested to participate in a confidential survey. Interviews were conducted to identify the current obstacles each faced with respect to efficient and effective data collection and transfer, processing, and reporting of environmental mercury levels. Some researchers provided copies of blank data collection sheets and field data recording protocols.

Phase 2. Software development

This phase focused on 2 general components: (1) identifying key design elements for developing a prototype mobile field data collection application specifically aimed at capturing the type of data needed to solve problems identified during Phase 1 and replicate field data recording protocols, and

(2) selection of appropriate hardware compatible with developing and testing the prototype software to meet overall project objectives. Data input screens for rapid capture of field data and geographic coordinates were developed using "C" programming language for mobile devices running the Palm Operating System (Table 1). The mobile application used 41 different input screens for electronic recording of data on loons, crayfish, prey fish and environmental data, such as water quality and lake information. This application was designed to support Global Positioning Systems (GPS) data acquisition via a custom interface used to record latitude/longitude coordinates from Magellan Companions and Garmin GPS devices. Different Portable Data Assistants (PDAs) were considered in light of availability, cost, user-friendliness, potential efficiency, and system compatibility (Table 1). (Fig. 1)

Phase 3. Product evaluation

This phase consisted of testing and evaluation of the resulting prototype software application. Six PDAs (Handspring Visors and Palm Pilots) were deployed in the summer of 2002 and a first generation mobile application was demonstrated to field biologists at BioDiversity Research Institute (BRI) for immediate feedback on the design and utility of the program. Limitations of the first prototype were discussed along with potential solutions and enhancements, leading to a re-design of the mobile application. Another custom-designed software application was developed using the "Delphi" programming environment to implement data transfer capabilities from the rudimentary file structure of the revised mobile application to a more robust database application operating on personal computers (PCs). The PC database was developed using Microsoft Access.

Phase 4. Product re-evaluation and re-engineering

This phase began with deployment of the re-designed prototype to field biologists at BRI and the Massachusetts Department of Environmental Protection (MA DEP) for testing and evaluation.

After reviewing evaluations, re-engineering decisions were made based on specific needs identified by the field biologists. At this juncture, a commercial software development company was retained, and modifications were planned after careful review of the results of earlier proof-of-concept designs. This iteration of the mobile field data collection module focused solely on expanding and refining data collection for the common loon (*Gavia immer*). The module was redesigned by experienced and specialized programmers skilled in mobile field data applications development, using an existing off-the-shelf application development environment (Pumatech Satellite Forms, now called Intellisync MobileApp Designer), Microsoft Access as the back-end database on a local workstation in conjunction with a data synchronization engine (Pumatech Enterprise Intellisync Server) (Table 1). Data input screens were designed to record specific observations regarding territorial pairs of loons, including nesting success and location, behavior and activity, and territory fidelity (Fig. 1). In addition, data input screens for color-coded leg bands provided for a rapid method of identifying loons and aid in understanding and analyzing the population dynamics (Fig. 1). Abiotic data, specifically water, was used to develop the integrated database model that included embedded GIS/mapping functionality using ESRI MapObjects Lite. The loon module was deployed for testing by field biologists from BRI and the Virginia Polytechnic Institute and State University (VT) and used to collect territory information regarding loon pairs in the Rangeley Lakes region of Maine.

Results and discussion

PDAs have been used for a number of different purposes, including the presentation of medical records and reports, patient encounters, accessing drug information, evaluating medical courses and reporting other observations (Buchauer et al., 1998; Speedie et al., 2001; McCreadie et al., 2002; Kwinter, 2003). Although other applications designed for mobile computing have proved successful, associated hardware, software and development costs continue to be a disadvantage for more widespread use (Buchauer et al., 1998; McCreadie et al., 2002; Speedie et al., 2001; Tschopp et al., 2002). In the present study, the development of a mobile handheld computer application that connects multiple field data

Table 1. Two separate software development strategies were used for the Mercury Tracker™ prototypes

Products	Components	Design strategy	Hardware used	Programming
Mercury Tracker™ 1.0 – Mobile Application	Mobile application	Mobile field data collection application specifically aimed at capturing data in a meaningful format and replicate field data recording protocols	Portable Data Assistant using Palm Operating System (Palm Vx, m105 and m130; Handspring Visor Neo and Visor Prism)	Customized programming for mobile application done in "C" programming language
Mercury Tracker™ 1.1 – Mobile Application	Custom synchronization conduit	Automating data transfer into user databases	GPS devices (Magellan Companion and Garmin)	"Delphi" programming environment used to implement data transfer capabilities from the mobile application file structure to a more robust database application operating on personal computers (PC's)
Mercury Tracker™ 1.1 – Data Transfer Conduit Mercury Tracker™ 1.1 – Desktop Database	Microsoft Access Databases		PC Laptop (IBM and Dell)	Microsoft Access
Mercury Tracker™ 2.0 – Loons Mobile Application	Mobile application with local data editing capabilities focused on loon data	Re-engineering of the original software to develop an integrated database environment with improved methods for aggregating and sharing data	Portable Data Assistant using Palm Operating System (Handspring Visor Prism)	Pumatech Satellite Forms
Mercury Tracker™ 2.0 – Loons Personal Database	Synchronization conduit for deposition of multi-source data directly into database		GPS devices (Magellan Companion)	Pumatech Enterprise Intellisync Server
Mercury Tracker™ 2.0 – Loons Integrated Database and Reporting System	Microsoft Access Database with embedded GIS interface		PC Laptop (Dell)	Microsoft Access ESRI MapObjects Lite

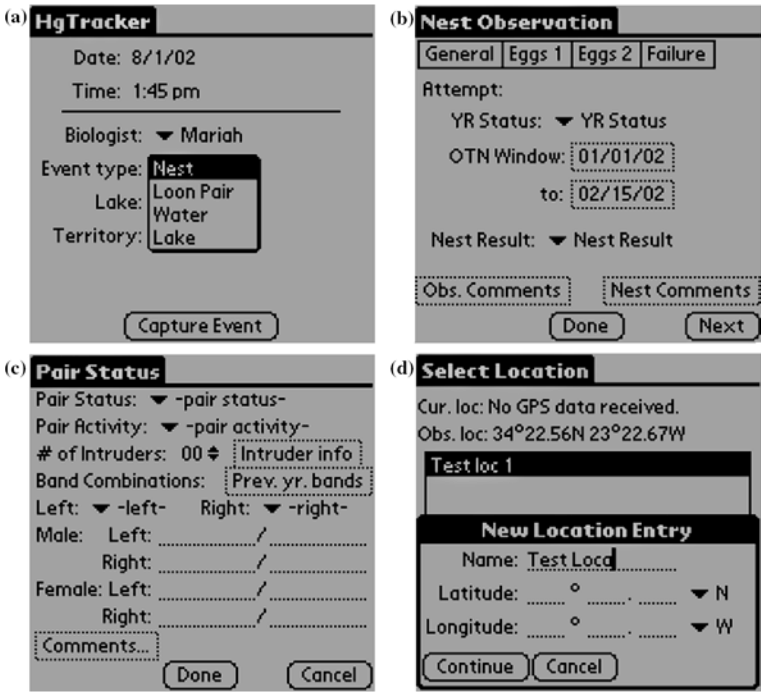

Figure 1. Sample PDA data input screens illustrating data capture features for *Mercury Tracker*™ 2.0 – *Loons Mobile Application.* (a) Drop down menu for event type selection. (b) Screen for nest observations. (c) Screen for loon pair status, including leg band identification. (d) Screen for GPS coordinates.

collection units to a centralized database standardized and automated the process of data collection, aggregation and sharing. The *Mercury Tracker*™ prototypes showed, using mercury as a model, that integrating complex assessments of environmental contaminants is possible and it is precisely the tool needed by collaborating organizations. This observation is based on the direct experience of the researchers involved in these Ecotoxicology issues where the integration of multiple disparate data sources on mercury was achieved only at a large time expense to those who had to integrate them.

The first generation software application, *Mercury Tracker*™ 1.0 focused on the collection and aggregation of data for tracking environmental indicators of mercury contamination in biota, namely crayfish, prey fish and common loons. Product evaluation through virtual field trials provided feedback on the efficiency and effectiveness of using an electronic infrastructure for capturing field data, including latitude and longitude coordinates. However, *Mercury Tracker*™ 1.0 data input forms proved to be cumbersome and information flow was inconsistent with general field operating procedures despite earlier detailed communications with collaborators.

The following three general areas were identified as being problematic to collaborators for integrating assessments of environmental mercury contaminants: (1) problems with differing individual and organizational data collection and management techniques; (2) problems associated with integrating multi-source data using the same parameters, for example, combining units of measurement (metric versus non-metric system), weight types (wet and/or dry) and geographic coordinates (decimal degrees versus latitude/longitude); and (3) problems associated with sharing and integrating multi-source data on the same and/or different parameters (biotic versus abiotic).

Phase 3 evaluations indicated that *Mercury Tracker*™ 1.1 was not as useful or effective as needed to achieve end user acceptance. Five major reasons for this were: (1) even though ensuring data integrity was identified as a key component of the prototype, the data could not be easily accessed and edited in the mobile application when legitimate modification of mobile records was required; (2) data in the mobile files could not be retrieved after initial entry for use on other screens and therefore background reference data entry had to be repeatedly entered for each event type; (3) the initial back-end database was not capable of managing multi-source, multi-disciplinary integrated assessments in a meaningful format; (4) although embedding a mapping and spatial display capability was desirable, it was not practical using the custom application programming techniques employed for the *Mercury Tracker*™ 1.1 series product; and (5) the prototype was not flexible enough to be modified by individual users as needed.

These findings led to a re-engineering phase using loons as the biota of choice for all of the *Mercury Tracker*™ components, resulting in the development of the following *cyberinfrastructure*: (1) *Mercury Tracker*™ 2.0 – Loons Mobile Application – a mobile field data acquisition system with local data editing capabilities; (2) *Mercury Tracker*™ 2.0 – *Loons Personal Database* – for synchronization of multi-source data into a meaningful format, and (3) *Mercury Tracker*™ 2.0 – *Loons Integrated Database and Reporting System* which included an embedded GIS/mapping interface (Table 1). Seamless downloading of data onto a laptop computer was performed every other day, with synchronization of new survey data from the mobile application into a Microsoft Access Database. This automated integration function of Loons Personal Database eliminated the need to transcribe handwritten notes from paper into a relational database, thus ensuring integrity of the data and increasing the efficiency of field data recording and use. For any given loon territory, field observations, such as the number of loons observed, the number of intruding loons observed, the activities of the loons in the territory, and nesting status could be easily and quickly extracted from the database. The *Loons Integrated Database and Reporting System* enabled aggregation and integration of incoming data from multiple mobile PDA devices and the reporting of observations over time and space.

User evaluations and testing conducted by field biologists during various stages of the project provided both proof-of-concept and proof-of technologies feedback and proved to be critical to the re-engineering cycles and to the overall success of the project. There was measurable success with incorporating data synchronization functionality needed to link mobile PDA devices to a central field computer, illustrating the ability to quickly and easily store and retrieve data collected by multiple teams of field researchers. The development team was also able to demonstrate how multiple field inspection oriented databases could be linked in order to aggregate data into a standardized data repository with an embedded GIS/map display supporting spatial views and analyses of that data. This type of capability has significant importance for field biologists tracking and reporting on environmental indicators of mercury contamination. The final application, *Mercury Tracker*™ 2.0, illustrated both the efficient and effective use of technology in the hands of field biologists for real-time data collection, transfer and integration, and demonstrated the relevance and utility of using an integrated, end-to-end all-digital approach. Data collection was accomplished easily and efficiently through the use of the drop down menus, check boxes, and automated GPS functionality, which ensured the use of standardized nomenclature and units of measure thereby facilitating data aggregation and sharing. In hindsight, working earlier on with commercial-off-the-shelf technology such as that employed for *Mercury Tracker*™ 2.0 during Phase 4, rather than embarking on a total customized programming effort at the outset, would most likely have led to being able to spend extra time and resources on a more integrated design and data exchange methodology.

Our approach was a general one that can be applied to many other types of environmental toxicology studies and field assessments in order to provide scientists with an advanced tool for recording and displaying data. It simplifies the overall process of collecting and merging data in an immediately useful and standardized format, consistent with the findings of Barthell et al. (2002) and Greenes et al. (2001). Furthermore, it minimizes the need for extensive knowledge of Geographic Information Systems (GIS) or associated third party software to conduct basic spatial queries of data, thereby facilitating real-time analyses of incoming data and laying the groundwork for a more extensive data-reporting framework. As such, our *cyberinfrastructure* serves to catalyze, support and enhance research collaborations on environmental mercury contamination being conducted by multiple disparate field research teams all working towards a common purpose, as represented in this issue of Ecotoxicology by the efforts of the Northeastern Ecosystem Research Cooperative (NERC). Architecture supporting the overall desired outcomes and goals of the project should be considered a necessary ongoing exercise in project planning, design and implementation phases. Integration of multi-source, multi-disciplinary data into a single database, and sharing of data between similar databases is a complex task that will likely change over time, necessitating a flexible design approach able to accommodate such changes as quickly and easily as possible.

Although all of the component parts needed to make this type of data collection and sharing arrangement possible have been shown to work and are viable at some level as a result of the *Mercury Tracker*™ prototype development effort, we believe our current application can be improved further and be more widely applicable by using a more flexible database architecture, thereby making it adaptable to changing scientific needs as determined by individual researchers, including an underlying architecture suitable for virtually any environmental pollutant. This architecture would address many of the challenges faced by the NERC researchers during the process of integrating databases for these Ecotoxicology issues.

Acknowledgements

The authors wish to thank the Dr C. Mark Smith and Mr Bob Maietta from the Massachusetts Department of Environmental Protection (MA DEP) and the Northeastern Ecosystem Research Cooperative (NERC). Financial support was provided by the National Science Foundation (NSF), Award Numbers 0196295 and 0238069. Visit NSF home page at http://www.nsf.gov; award abstract http://www.nsf.gov/home/crssprgm/powre/powre.htm.

References

Bajcsy, R. (2001). NSF Blue Ribbon Panel on Cyberinfrastructure, Advisory Committee for Cyberinfrastructure.

Barthell, E.N., Cordell, W.H., Moorhead, J.C., Handler, J., Feied, C., Smith, M.S., Cochrane, D.G., Felton, C.W. and Collins, M.A. (2002). The Frontiers of Medicine Project: a proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. *Ann. Emerg. Med.* **39**, 422–9.

Buchauer, A., Werner, R. and Haux, R. (1998). Cooperative problem Solving with Personal Mobile Information Tools in Hospitals. *Methods Inform. Med.* **37**, 8–15.

Evers, D.C. and Clair, T.A. (2005). Mercury in Northeastern North America: A synthesis of Existing Databases. *Ecotoxicology* **14**, 7–14.

Evers, D., Kaplan, J., Meyer, M., Reaman, P., Braselton, W., Major, A., Burgess, N. and Scheuhammer, A. (1998). A geographic trend in mercury measured in common loon feather and blood. *Environ. Toxicol. Chem.* **17**, 173–83.

Greenes, R.A., Peleg, M., Boxwala, A., Samson, T., Patel, V. and Shortliffe, E.H. (2001). Sharable computer-based clinical practice guidelines: rationale, obstacles, approaches, and prospects. *Medinfo.* **10**, 201–5.

Kamman, N.C., Loray, P.M., Driscoll, C.T., Estabrook, R., Major, A., Pientka, B. and Glassford, E. (2004). Assessment of mercury in waters, sediments and biota of New Hampshire and Vermont Lakes, USA, sampled using a geographic randomized design. *Environ. Toxicol. Chem.* **23**, 1172–86.

Kwinter, S. (2003). Trailblazer. Wired 06: 144.

McCreadie, S.R., Stevenson, J.G., Sweet, B.V. and Kramar, M. (2002). Using personal digital assistants to access drug information. *Am. J. Health-Syst. Pharm.* **59**, 1340–3.

NESCAUM (1998). *Northeast States and Eastern Canadian Provinces Mercury Study*. NESCAUM, NEWMOA, NEIWPCC, EMAN, Boston, MA.

Speedie, S., Pacala, J., Vercellotti, G., Harris, I. and Zhou, X. (2001). PDA support for outpatient clinical clerkships: mobile computing for medical education. *Proceedings American Medical Informatics Association Symposium*. 632–6.

Stein, L. (2003). Integrating Biological Databases. *Nature Review Genetics*. **4**, 337–45.

Tschopp, M., Lovis, C. and Geissbuhler, A. (2002). Understanding usage patterns of handheld computer in clinical practice. *Proceedings American Medical Informatics Association Symposium*. 806–9.

Wiener, J.G., Krabbenhoft, D.P., Heinz, G.H. and Scheuhammer, A.M. (2003). Ecotoxicology of mercury. In D.J. Hoffman, B.A. Rattner, G.A. Burton Jr. and J. Cairns Jr. (eds). *Handbook of ecotoxicology*, pp. 409–463. Boca Raton, FL: Lewis Publishers.

Wolfe, M.F., Schwarzbach, S. and Sulaiman, R.A. (1998). Effects of mercury on wildlife: a comprehensive review. *Environ. Toxicol. Chem.* **17**, 146–60