# An Automated, Web-Enabled and Searchable Database System for Archiving Electrogram and Related Data from Implantable Cardioverter Defibrillators

W Zong[1,2], P Wang[3], B Leung[2], GB Moody[2], RG Mark[2]

[1] Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[2] Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA
[3] New England Medical Center, Tufts University Medical School, Boston, MA, USA

## Abstract

*The advent of implantable cardioverter defibrillators (ICDs) has resulted in significant reductions in mortality in patients at high risk for sudden cardiac death. Extensive related basic research and clinical investigation continue. ICDs typically record intra-cardiac electrograms and inter-beat intervals along with device settings during episodes of device delivery of therapy. Researchers wishing to study these data further have until now been limited to viewing paper plots. In support of multi-center clinical studies of patients with ICDs, we have developed a web based searchable ICD data archiving system, which allows users to use a web browser to upload ICD data from diskettes to a server where the data are automatically processed and archived. Users can view and download the archived ICD data directly via the web. The entire system is built from open source software. At present more than 500 patient ICD data sets have been uploaded to and archived in the system. This project will be of value not only to those who wish to conduct research using ICD data, but also to clinicians who need to archive and review ICD data collected from their patients.*

## 1. Introduction

Sudden cardiac death due to coronary artery disease and/or ventricular arrhythmias is a leading cause of death in the United States. Nearly 250,000 out-of-hospital cardiac arrests occur annually in the US and more than 95 percent of victims of out-of-hospital cardiac arrest die [1]. Researchers have found that the death rate in high-risk heart attack patients can be substantially reduced by implantable cardioverter defibrillators (ICDs) [2]-[4].

It is estimated that more than 150,000 patients worldwide are now using ICDs and nearly 50,000 new devices were implanted in the past year. The number of new ICD implants is increasing at a rate of more than 20 percent a year [5].

ICDs save lives. The technology is expensive, however, and not without some risks such as inappropriate shocks, surgical and device-related complications, and worsening of heart rhythm. Extensive basic research and clinical investigations continue to determine optimal criteria for selection of suitable patients, to improve algorithm performance and cost effectiveness, and to minimize risks [1]. Furthermore, the data acquired from ICDs supports clinical studies of basic cardiac electrophysiology, anti-arrhythmic therapeutics, and environmental and behavioral triggers of ventricular arrhythmias [6].

ICDs typically record intra-cardiac electrograms (EGMs) and inter-beat intervals during periods immediately preceding and following device therapy, in order to document the life-threatening arrhythmias they detect and the effects of the therapeutic actions they apply. ICDs also document programmable ICD settings and cumulative device-related or physiological measurements during use. ICD manufacturers provide proprietary hardware and software "programmers" that can interrogate ICDs in order to retrieve these data and store them on floppy disks.
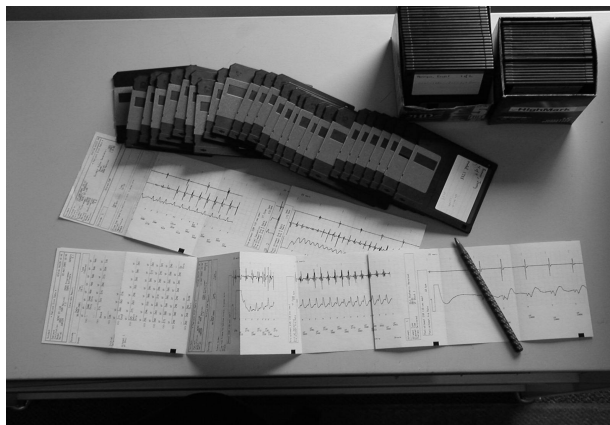


Figure 1. The existing situation for viewing ICD data

Researchers wishing to study these data further have until now been limited to viewing paper plots that can only be made by the proprietary "programmers" (figure 1). The data are generally not available in usable computer-readable form. Another problem is that floppy disks are inconvenient to index and are not reliable storage media. These limitations loom as a barrier to efficient investigations on large volumes of collected ICD data.

In support of NIH-funded multi-center studies, we have developed an automated, web-based and searchable ICD data archiving system. The system allows clinicians at each center to transfer data from the original ICD diskettes to a local "client system" that indexes and archives the original data. In addition, the client system de-identifies the data and uploads them to an Internet-based archive shared by all participating centers. There the data are automatically reformatted and stored in a relational database. Using the shared archive, clinicians and researchers can view the EGM waveforms and related ICD data online via the Web, search the database, add clinical annotations and comments, and download data at will. All the data are in a standard (PhysioNet) format [7, 8]. Using this system, researchers can efficiently conduct basic and clinical studies on large-scale collections of ICD data. The system is particularly suitable for supporting multi-center studies on patients with ICDs since investigators can use the archive at any location where Internet access is available.

## 2.    Methods

The ICD data archiving system consists of two parts: the client systems (generally located in clinical facilities) and the Internet-based shared archive. (See figure 2.)
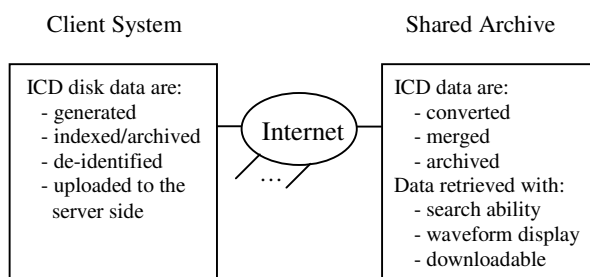


Figure 2.  System overview diagram

At each client site, the original ICD disks are uploaded to an internal server where the original disk images are properly indexed, archived, and de-identified. The de-identified disk images are then uploaded to the Internet-based shared archive where the data are automatically converted to PhysioNet format, merged (to remove redundant data among multiple ICD disks), and inserted into a relational database. The shared archive provides search capability, online waveform display, and downloadable data in PhysioNet format.

The shared archive's data format is that used by PhysioBank [7], permitting researchers to read and analyze the data using a wide variety of software available freely from PhysioToolkit [7]; since the format is fully documented, researchers can created custom software for these purposes if desired.

Each client system archives ICD data disk images in their original (proprietary) formats, so that it is always possible for investigators to recreate copies of their own original disks for analysis or review using the proprietary "programmers" if necessary.

## 2.1.    Client system

The client systems (figure 3) are designed around an internal Linux PC that performs the tasks of ICD disk indexing/archiving and de-identification. The Linux PC runs a web server and provides the internal networking function. The customised software for receiving, indexing, archiving, and de-identifying ICD data runs on the Linux PC.

One or more Windows PCs can be connected to the server machine via the internal network, working as client machines running web browsers. The web server provides pages to guide users in uploading or retrieving ICD disk data.
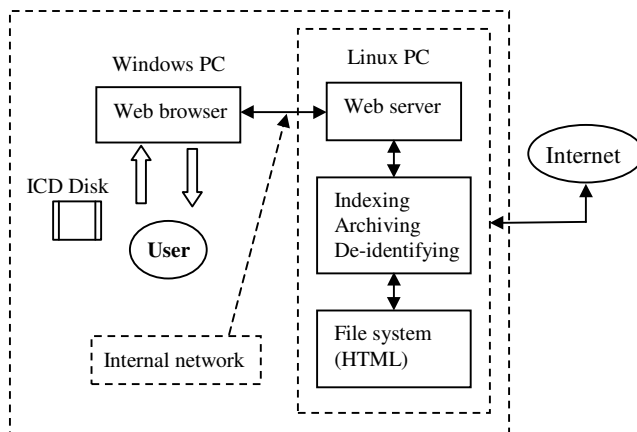


Figure 3. The client site ICD disk index/archive system

Uploading diskettes is guided by an upload utility form that requests patient identification and upload authority codes. Upon receiving a disk upload from the client machine, the internal Linux PC activates programs to index and store the disk data in a locally accessible archive. The system also supports data retrieval from the local archive using a web browser. It is convenient to browse the local archive to find particular ICD disks since all data are indexed with patient name, study number, etc., and all information is organized in HTML.

The internal network-based client system permits multiple clinical users to upload and access the disk data simultaneously.

In addition to its indexing and archiving functions, the client-based software also generates a de-identified version of the disk data to protect patient confidentiality. The anonymized disk images are then automatically or manually uploaded to the shared archive.

## 2.2. Shared archive

A Linux PC running a web server and a relational database server hosts the shared archive. A large suite of customised software for converting, inserting, and displaying ICD data was developed for this system.

The structure of the system is shown in figure 4. The web server functions as an interface through which the client systems upload anonymized ICD data to the system, and access EGM and related data stored in the system.

When the shared archive receives an ICD disk image from a client system, customised processing software converts the EGM and annotations into PhysioNet format and inserts the accompanying data into a relational database. A middle-ware layer facilitates inserting the data into the PostgreSQL relational database.
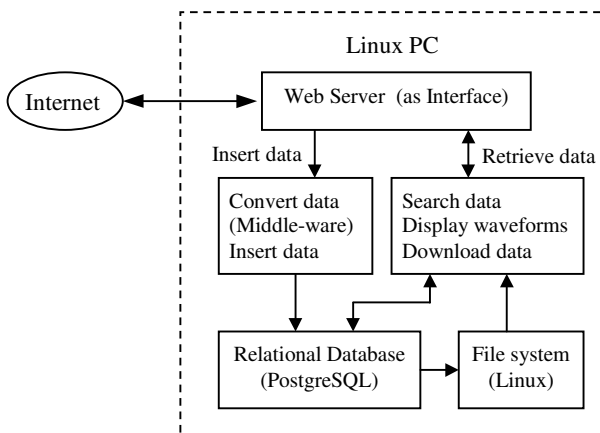


Figure 4. The shared ICD data archiving system

Retrieval of data from the database is supported by interactive web pages that prompt the user to enter and select search criteria. The search criteria are translated into SQL queries and sent to the relational database. The query results are then used to create web pages sent to the user's browser for display. The system supports online EGM waveform display with an in/out zoom feature. The EGM digital signal data can be downloaded in PhysioNet format at will. Other types of ICD data, such as device settings and therapy mode, can be viewed online and downloaded as well.

## 2.3. Open-source software

The design and implementation of the system were based on open source software, including the Linux operating system, the Apache web (HTTP) server, and the PostgreSQL relational database. All the customised utility software (such as web interfacing, data conversion and insertion, and waveform display) was developed using Perl, Perl CGI, C, SQL and the WFDB software package in PhysioToolkit [7] (the open source software repository of PhysioNet, the NIH-funded Research Resource for Complex Physiologic Signals [8]).

These design and implementation choices allow maximum flexibility for investigators and developers. Use of open source products substantially reduces the costs of development and operation of the system, and increases its reliability.

## 3. Results

The first version of the complete system has been implemented, and is capable of processing ICD data from all (approximately 45) models of Guidant ICD devices currently in active use. Future versions of the system will accept ICD data from other manufacturers such as Medtronic and St. Jude.

The current system has been operating since April 2002 with a client site at the New England Medical Center and the shared archive at the Massachusetts Institute of Technology (MIT). As of September 2002, ICD data from more than 500 patients using Guidant devices had been uploaded and archived.

The system is currently supporting two multi-center clinical studies (the NIH-sponsored TOVA study [6], and a follow-up to the MADIT II [4] study), as well as clinical cardiology services at New England Medical Center.
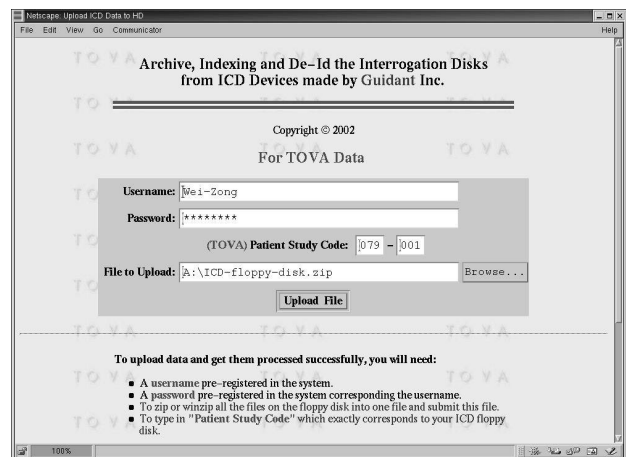


Figure 5. The web interface for uploading of an ICD disk image to the client system

Clinicians find the system easy to use. A familiar web browser interface supports data uploading, viewing or downloading. Figure 5 shows the web interface for uploading of the original ICD disk data to the client system. Users need only enter proper authorization codes, the study code, and the name of the file to upload. Clicking the "Upload File" button initiates the uploading and processing of the data.

Users can conveniently view all archived data online, including EGM waveforms, beat interval annotations, device settings, therapy modes, etc. The data may also be downloaded from the archive in PhysioNet format for further computerized analysis.

Figure 6 is a screen shot of an EGM episode from an ICD data record archived in the system. The beat and other annotations produced by the ICD device are displayed along with the EGM waveform. The waveforms with annotations can be zoomed in and out for optimal examination.
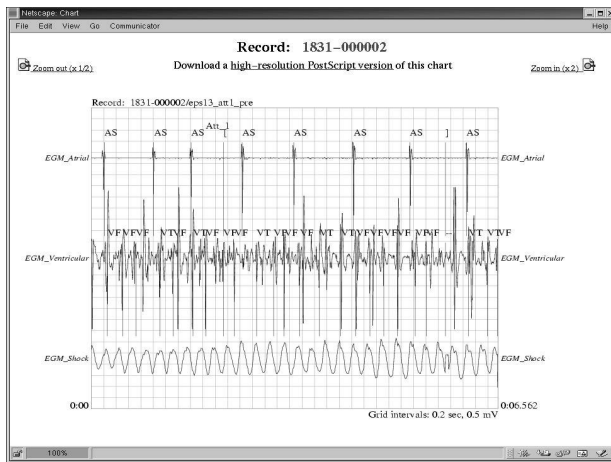


Figure 6. A screen shot of EGM waveform display for an ICD data record archived in the system

Users may search the archive's database applying a variety of search criteria, including patient age, gender, implantation date, ICD manufacturer, device model, cardiac rhythm type, therapy type, etc. The system also supports the attachment of arbitrary clinical notes/annotations to EGM episodes. Those notes are also searchable.

## 4.    Conclusion and discussion

An automated, web-based, and searchable ICD data archiving system has been developed for the first time. It is designed to accept ICD data from multiple manufactures and models. It provides local archiving at the client site, and Internet-based archiving of de-identified patient data from multiple sites. The system provides a permanent searchable repository of valuable ICD data in accessible digital format, and is ideally suited to support multi-center clinical trials as well as basic research on ICD function, algorithm design, and the electrophysiology of sudden cardiac death.

This database system will significantly enhance the efficiency of multi-center clinical investigations of ICD technology and basic cardiac electrophysiology.

Further work will expand the current system to include ICD data from devices made by Medtronic, Inc. and St. Jude, Inc., the other two major ICD manufacturers.

## Acknowledgements

## References

[1] Bigger JT. Expanding indications for implantable cardiac defibrillators. N Engl J Med 2002; 346:931-933.

[2] Moss A, Hall W, Cannom D, et al. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. N Engl J Med 1996; 335:1933-1940.

[3] Buxton A, Lee K, Fisher J, et al. A randomized study of the prevention of sudden cardiac death in patients with coronary artery disease. N Engl J Med 1999; 341:1882-1890. [Erratum, N Engl J Med 2000; 342:1937-1945.]

[4] Moss A, Zareba W, Hall W, et al, for the Multi-center Automatic Defibrillator Implantation Trial II Investigators. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. N Engl J Med. 2002; 346:877-883.

[5] http://www.rochester.edu/pr/Currents/V29/V29N16/ story06.html

[6] http://www.clinicaltrials.gov/ct/gui/show/NCT00005243

[7] http://www.physionet.org/

[8] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 2000; 101(23):e215-e220.

Address for correspondence:

Wei Zong, Ph.D.
Room E25-505, MIT
77 Massachusetts Avenue
Cambridge, MA 02139, USA
E-mail: wzong@mit.edu