

Bruce Bartholow Duncan^I

Álvaro Vigo^I

Émerson Hernandez^I

Vivian Cristine Luft^I

Hubert Ahlert^{II}

Kaiser Bergmann^{III}

Eduardo Mota^{IV}

Information Management in Multicenter Studies: the Brazilian Longitudinal Study for Adult Health

ABSTRACT

Information management in large multicenter studies requires a specialized approach. The *Estudo Longitudinal da Saúde do Adulto* (ELSA-Brasil – Brazilian Longitudinal Study for Adult Health) has created a Datacenter to enter and manage its data system. The aim of this paper is to describe the steps involved, including the information entry, transmission and management methods. A web system was developed in order to allow, in a safe and confidential way, online data entry, checking and editing, as well as the incorporation of data collected on paper. Additionally, a Picture Archiving and Communication System was implemented and customized for echocardiography and retinography. It stores the images received from the Investigation Centers and makes them available at the Reading Centers. Finally, data extraction and cleaning processes were developed to create databases in formats that enable analyses in multiple statistical packages.

DESCRIPTORS: Multicenter Studies as Topic, methods. Databases as Topic. Database Management Systems. Information Systems. Information Storage and Retrieval. Cohort Studies.

^I Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

^{II} Centro de Processamento de Dados, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

^{III} Faculdade da Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

^{IV} Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, BA, Brasil

Correspondence:

Bruce B. Duncan
PPG-Epidemiologia
R. Ramiro Barcelos, 2600/414
90035-003 Porto Alegre, RS, Brasil
E-mail: bbduncan@ufrgs.br

Received: 10/5/2011

Approved: 6/5/2012

INTRODUCTION

In 2001, Guimarães et al⁷ described that some areas of scientific activity in epidemiology, like that of chronic non-communicable diseases, were still not well developed in Brazil. In the last decade, as a result of a substantial increase in the promotion of health research, these diseases have begun to be better investigated. The comprehensiveness of the collected information has been expanded and the methodology for its acquisition has become more complex.

One example of the new scenario of Brazilian epidemiological research is the creation of a network of investigators for the implementation of the *Estudo Longitudinal de Saúde do Adulto* (ELSA-Brasil – Brazilian Longitudinal Study for Adult Health),¹ a cohort study that aims to follow 15,000 adults over time at six Brazilian centers. The evaluations involve interviews and tests of varied complexity, including images from echocardiography and retinography. Aspects like the number of participants and of interviews and tests, the longitudinal nature, multicenter organization, and the complex assessments of ELSA determined the characteristics of its data system.

At the end of the 1970s, one of the first epidemiological studies on chronic diseases carried out in Brazil included adults in a probability sample from the State of Rio Grande do Sul (Southern Brazil). The information was obtained by means of home interviews and the measurement of blood pressure, and entered on paper to be subsequently keyed and entered into a mainframe computer, a procedure that was typical of that time.⁴ A similar methodology was employed in the following decade in the Study of Diabetes Prevalence, a multicenter study that was conducted in many Brazilian states, with data keying centralized in São Paulo.^{9,a} Another study in that decade, the Study of Risk Factors for Chronic Non-communicable Diseases, entered its information on paper, and data was entered using microcomputers.⁵ In the 1990s, a multicenter study about diabetes in pregnancy used the EpiInfo software to check for inconsistencies during double keying, to control skip errors in the interviews, and to detect incongruent values in the measurements.¹¹

Entering information on paper in these studies required revision and coding by a project supervisor before keying, which many times occurred when it was no longer possible to correct the detected errors. To reduce this problem, the Study on Consumption and Eating Behavior in Pregnancy used a system that eliminated the keying stage, controlled skip errors in the interviews and detected incongruent values. The method was to scan the questionnaires and forms so that the data could

be captured by the program Teleform (Cardiff, Vista CA USA). The procedure did not eliminate the need to use paper to enter the information and, when the scanning was delayed, the potential for error control was limited. Some randomized clinical trials, whose volume of collected information from each individual is usually low, have started to use web systems for data entry and management.²

Table 1 highlights the evolution of the data collection and entry procedures in these studies and in some classic international epidemiological studies of chronic diseases over the same period. The trend is to use systems entering data into microcomputers, which ensures better data confidentiality and allows immediate detection and correction of errors, often while the individual is still with the field team.⁶

When data collection is performed in the environment of a Research Center, the tendency is to enter the data via the internet, connected with an information management system.^{10,13} Besides allowing greater data reliability, such systems can call attention to the presence of situations of interest (for example, diabetes) and can direct specific situations according to the characteristics of the research individuals (for example, if an individual has previously undergone bariatric surgery, he should not be submitted to a glucose tolerance test). With a continual and effective monitoring of the data and immediate correction, these systems also enable data to be quickly transformed into analysis databases.

The ELSA system was designed to incorporate the advantages of web systems. This paper aims to show the structure of the ELSA's Datacenter and to describe the methods of information entry, transmission and management in the Study.

DATACENTER

ELSA's Datacenter is composed of a multidisciplinary team from the areas of epidemiology, statistics, information technology (analysts, programmers and a web designer) and public relations. As one of the aims of the project is the enhancement of research capabilities, the Datacenter offers the possibility of participation of undergraduate and postgraduate students, grouping them into three distinct teams (Figure 1): Biostatistics and Epidemiology, Statistical Programming, and Systems. The Center has consultants from the *Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul* (CPD-UFRGS – Data Processing Center) for technical

^a Centers for Disease Control and Prevention. EpiInfo Version 6: a word processing, database and statistics program for public health on IBM-compatible microcomputers [computer program]. Atlanta; 1995.

Table 1. Data collection and entry systems in some Brazilian and international epidemiological studies about chronic diseases.

Study	Year	n	Data collection and entry system
Brazilian			
Salt and Blood Pressure in Rio Grande do Sul, Brasil	1978	4,565	Paper + mainframe keying
Study of Risk Factors for Chronic Non-communicable Diseases – RS	1986	1157	Paper + microcomputer keying
Study of Diabetes Prevalence	1986	21,847	Paper + mainframe keying
BambuÍ Project	1996	1,606	Paper + microcomputer keying
Pro-Health Study	1999	4,030	Paper + microcomputer keying
ECCAGE	2006	2,590	Paper + scanning (Teleform)
ACT	2008	2,308	Web system
International			
ARIC (United States)	1987	15,792	Direct entry - microcomputers; subsequently, web system
CHS (United States)	1989	5,201	Direct entry - microcomputers; subsequently, web system
AUSDIAB (Australia)	1999	1,478	Paper + scanning (Teleform)
MESA (United States)	2000	6,814	Paper + scanning (Teleform); subsequently, web system
Kadoorie (China)	2004	500,000	Direct entry - microcomputers
BioBank (United Kingdom)	2007	500,000	Direct entry - microcomputers
HCHS/SOL (United States)	2008	16,000	Web system

ECCAGE: Study on Consumption and Eating Behavior in Pregnancy

ACT: Acetylcysteine for the Prevention of Contrast-Induced Nephropathy

ARIC: Atherosclerosis Risk in Communities Study

CHS: Cardiovascular Health Study

AUSDIAB: Australian Diabetes, Obesity, and Lifestyle Study

MESA: Multi-Ethnic Study of Atherosclerosis

HCHS/SOL: Hispanic Community Health Study – Study of Latinos

support, as well as international experts (members of the team of the Collaborative Studies Coordinating Center of the University of North Carolina).

To support the activities of the Datacenter, ELSA created a System Development Nucleus, composed of representatives of the Investigation Centers (ICs) with experience in information technology and systems development, and of two members of the Datacenter itself.

The coordinator of the Datacenter is a member of the project's Steering Committee and participates in the general planning of the study. The activities of the Datacenter during the baseline focused on the creation of mechanisms for data entry, processing and cleaning, on receiving images from the Investigation Centers and on sending them to the Reading Centers, and on the creation and distribution of the study's databases. The operational link between the ICs and the Datacenter is performed by data managers at each center.

THE STUDY'S INFORMATION SYSTEMS

The information systems that were developed consist of web modules – the ELSA System – that provide acquisition forms for each study activity, and of a Picture Archiving and Communication System, PACS – PACS-ELSA.

One of the first decisions made in system selection was to use public domain software. Advantages concerning price, business model and support are some of the factors that influenced this decision. ELSA followed the instructions of the federal government of developing software in a free and open way.^b

The systems are physically hosted at the CPD-UFRGS, where virtual servers were created to meet the development, testing and production needs. For security reasons, a specific server was created to host the production database, which is accessible only via the computers of the UFRGS network. Communication of study servers with the other centers occurs via *Rede Nacional de Pesquisa* (RNP – the National Research Network),^c allowing adequate

^b Brasil. Governo Federal. Diretrizes de uso de Software Livre no Governo Federal. Brasília (DF); 2011 [cited 2011 Mar 30]. Available from: <http://www.softwarelivre.gov.br/planejamento-cisl/diretrizes>

^c Rede Nacional de Pesquisa. Relatório de gestão RNP: edição anual 2012. Rio de Janeiro; 2012 [cited 2013 Mar 04]. Available from: http://www.rnp.br/rnp/relatorio_gestao.html 201

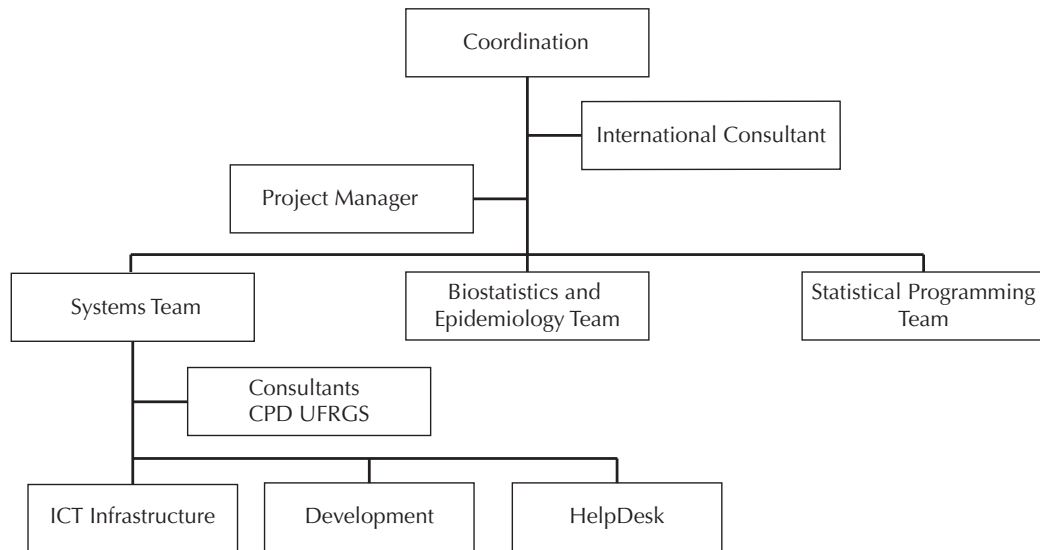


Figure 1. Structure of the ELSA's Datacenter (CPD-UFRGS: Data Processing Center of Universidade Federal do Rio Grande do Sul; ICT: Information and Communication Technology).

transmission speed, band width and availability. The PACS functions similarly in a virtual server.

THE ELSA SYSTEM

Development Options and Characteristics of the ELSA System

As mentioned, one of the first decisions in the development of the ELSA System was to use public domain software, instead of paid tools and technologies such as C# of Microsoft®. Another initial decision, considering the longitudinal character of the research and the diversity in the origin of the data, was not to use desktop/offline software like Epi6 and client-server systems (a well known example of which is that used for income reporting for tax purposes in Brazil), as well as systems that capture images from paper forms (like the Cardiff Teleform) or other modalities listed on Table 1.

The adoption of the web system enabled automatic access not only to the validation rules for data entry, but also to the information already stored in the system (for example, preferential forms of contact with the participants). As a web system with the required features was not publicly available, a new system was created with support from the CPD-UFRGS. The Java platform was chosen as the development tool and execution environment, given the flexibility it offers. Table 2 presents a list with the main public domain softwares that were used.

In the ELSA System, the user can use tabs and menus to access the research forms, such as those for recruitment, reception, questionnaires and tests. There is also a keying interface for data collected initially on paper,

including auditing data (date, time and technician responsible for the collection). With online entry, the auditing information is automatically captured.

For baseline tests and interviews (Wave 1), 38 activity forms were created, all available online and on paper, the latter for use in case of technical problems (for example, collection in places outside of the research centers, moments of electrical failure or of problems in the internet network used to communicate with the central server). Aspects of the study's design, studied population, recruitment and baseline tests and interviews have been detailed in another publication.¹

Given that the ELSA System is an integrated system with information sharing at all locations of data entry, it offers management tools that aid in the communication between centers (for example, data obtained in the acquisition of specialized tests are automatically made available to the reading centers) and in the administration of data collection at the ICs (for example, changes in the information of staff using the system are also automatically made available).

System Development Process

The development process of the ELSA System was incremental, passing progressively through the stages of analysis, programming and testing. In this strategy, called iterative and incremental development, the software construction begins with a small prototype and, after successive refinements, the total system is constructed. The strategy was very useful in the case of ELSA allowing learning time in the initial stage of this novel process. In addition, it enabled those involved in the programming process to discover important details

Table 2. Main public domain softwares used for the development of the ELSA System.

Element	Technology	Available from
Application Server	JBoss	www.jboss.org/
Operating System	Debian	www.debian.org
Creation of Questionnaires	Opinio	objectplanet.com/opinio/
Development Environment	Eclipse	www.eclipse.org/
DMS	PostgreSQL	www.postgresql.org/
Programming Language	Java	www.java.com/

DMS: Database Management System

in earlier stages of development, making change and/or adaptation easier.⁴ Besides this incremental strategy, the development process also employed agile methods like Scrum for project management and planning¹² and XP (Extreme Programming) for software development.⁸ Programming was performed in iterations, usually of three or four weeks.

As the system was being developed, the necessity of having a production manager responsible for the development of a detailed “vision of the product” became apparent. This vision enabled conducting a more detailed initial discussion about the system’s general needs before beginning to program its parts. It helped identify interested individuals, in and out of the Datacenter, and outline predicted needs. Based on this vision, the Systems Team translated the needs into a proposal of system features. After each activity was reviewed and approved by the project manager and by the Datacenter coordinator (production manager), the Datacenter started the programming of the specific instrument.

Design of the Forms

The main steps of the process of construction of a collection instrument (form) in the system are summarized on Table 3. These forms were initially designed on paper by ELSA investigators, who were responsible for the corresponding instruments. The Datacenter team revised and standardized aspects like skip rules, treatment of missing values and minimum and maximum limits for variables, entering the process in a document called the map of variables. Rules for mandatory data entry were established. For questionnaires, data input was mandatory and minimum and maximum values rigidly fixed; for exams, entry of values outside minimum and maximum limits and missing entries generated alerts requiring confirmation of the unexpected (or missing) values in question before proceeding. After being structured, standardized and approved, the data entry instrument was sent to the Systems Team to translate its content into software specifications and for subsequent programming.

Table 3. Stages of the availability of features in the data system.

- The Biostatistics and Epidemiology Team contacts the researcher in charge of each instrument (form, questionnaire, report, etc.) to collect the necessary specifications for its implementation in the system;
- The Statistical Programming Team revises and standardizes the form, specifically aspects like skip rules and treatment of missing values and limits, generating maps of variables;
- The Datacenter coordinator establishes priorities for the implementation of new features in the system (based on the demands of the Steering Committee);
- The Systems Team translates the specifications contained in the maps into software specifications, and programs, performs initial tests, and makes the new features available for testing by the Biostatistics and Epidemiology Team;
- After testing the feature, the Biostatistics and Epidemiology Team contacts again the ELSA researcher responsible for the instrument to verify the acceptability and to certify the feature in the system;
- The Systems Team incorporates the new data entry and data extraction features in the production server;
- The Statistical Programming Team, following the instructions of the Biostatistics and Epidemiology Team, extracts, processes, and cleans the data, and develops data dictionaries and other documents to enable statistical analyses.

Once programmed, the form’s next stage involved the performance of internal tests, which were conducted first by the Systems Team (integration and exploratory tests) and then by the Biostatistics and Epidemiology Team (acceptance tests in a separate server used for testing and training). The integration and exploratory tests were performed right after programming. As soon as new features were considered to be ready by the Systems Team, they were made available for the acceptance tests. This second group of tests verified, at the end of each cycle, that the new features were in accordance with the project’s needs. These latter tests were recorded and saved for subsequent use by the Systems Team; obviating the need for the

⁴ Beck K, Beedle M, Bennekum A, Cockburn A, Cunningham W, Fowler M, et al. Manifesto para o desenvolvimento ágil de software. [cited 2011 Mar 2]. Available from: <http://manifestoagil.com.br/>

Biostatistics and Epidemiology Team to repeat these tests in the future.

After a new feature was approved internally, a user acceptance test is performed by the responsible ELSA investigator. Once certified as apt for use, the feature was integrated into the data entry system of the production server for use.

Use of the System

After the development and certification of the data collection instruments, the members of the ELSA team at the diverse ICs entered data directly online or by keying data from the corresponding paper form. A support system – a HelpDesk for users – was created to resolve doubts, receive reports and give feedback concerning potential programming bugs. The HelpDesk is available by telephone and by a specific e-mail, being managed by a public relations intern.

Extraction and Preparation of Bases for Analysis

The information collected via the ELSA System is stored in a PostgreSQL® database, with separated data tables for each activity. Extraction and transfer of these data to statistical analysis bases were periodically performed, usually on a monthly basis. Data extraction is carried out with the program SAS® (Statistical Analysis System), involving several stages (Figure 2). The data of each activity (module) are individually extracted, via SAS software, through a data visualization table produced by the Systems Team. The extractions are organized by date and the data of each activity are stored in a corresponding file. Throughout data entry, reports were made to inform ELSA investigators about the status of the process.

Monitoring of Data Entry

Monitoring and cleaning activities were coupled with the extraction process.⁶ To monitor the entry of each participant's data, a variable that indicated the performance of each activity was created, called a flag. When there are no data in the system for a given activity, its flag variable remains blank, thus allowing identification of participants with incomplete data in the system. When it is known that a participant will not perform one or more activities, a special value can be attributed to the indicator variable of the activity in question.

When joined together, the indicator variables (flags) form a file called the "Flag Report", which is stored in SAS format (sas7bdat) and.xls. The Flag Report in the.xls format is sent to the ICs for the identification of participants with incomplete data entry.

Data Cleaning

In the extraction, special values are attributed to identify specific types of non-responses, such as skips when the question or test does not apply, refusal to answer some question, etc. In this way, unexpected values, such as missing or inconsistent values, can be identified and communicated to the ICs for elucidation. A report on potential inconsistencies is created after the extraction and sent for revision at the ICs. The IC's data manager makes the possible corrections using the data entry web system and communicates the results of this revision to the Datacenter.

Generation of Databases for Analysis

When all the information has been entered into the system and the monitoring and cleaning stages are complete, the SAS database is "frozen" to start the study's formal analyses. Preliminary databases were periodically distributed to the IC coordinators to enable a preliminary examination of the data. In these databases, the participants are labeled with identification numbers that are different from those used in the collection process. The SAS files generated for analyses are distributed also in other formats, like SAV (SPSS) and DTA (Stata); thus, the investigators can conduct analyses in SAS, SPSS, Stata, R and other programs.

In addition, a "data book" was created for each activity, containing a succinct description of the information of each variable. For the quantitative variables, it presents the number of valid observations, minimum and maximum values, mean and standard-deviation. For the qualitative variables, it presents the frequencies of each category, number of valid observations and of missing values. Additionally, a system with graphical interface has been created to enable selection of desired variables to create a specific analysis database in a desired file format. This system can generate the database and the corresponding dictionary of variables, saving it in a previously specified location.

Security Aspects in the Use of the System

The Brazilian Society of Health Informatics^e presents a set of requisites, classified as obligatory and desirable, adapted from the International Standards Organization (ISO), which defines basic characteristics for the construction of an adequate electronic health registration system. ELSA implemented its data entry system considering it an electronic health registration system, meeting, whenever possible, the obligatory and desirable requisites, as described below:

^e Sociedade Brasileira de Informática em Saúde. Manual de requisitos de segurança, conteúdo e funcionalidades para sistemas de registro eletrônico em saúde (RES); versão 2.1. São Paulo: Conselho Federal de Medicina; 2004 [cited 2011 Mar 30]. Available from: http://www.uel.br/projetos/oicr/pages/arquivos/GTCERT_20040219_RT_V2.1.pdf

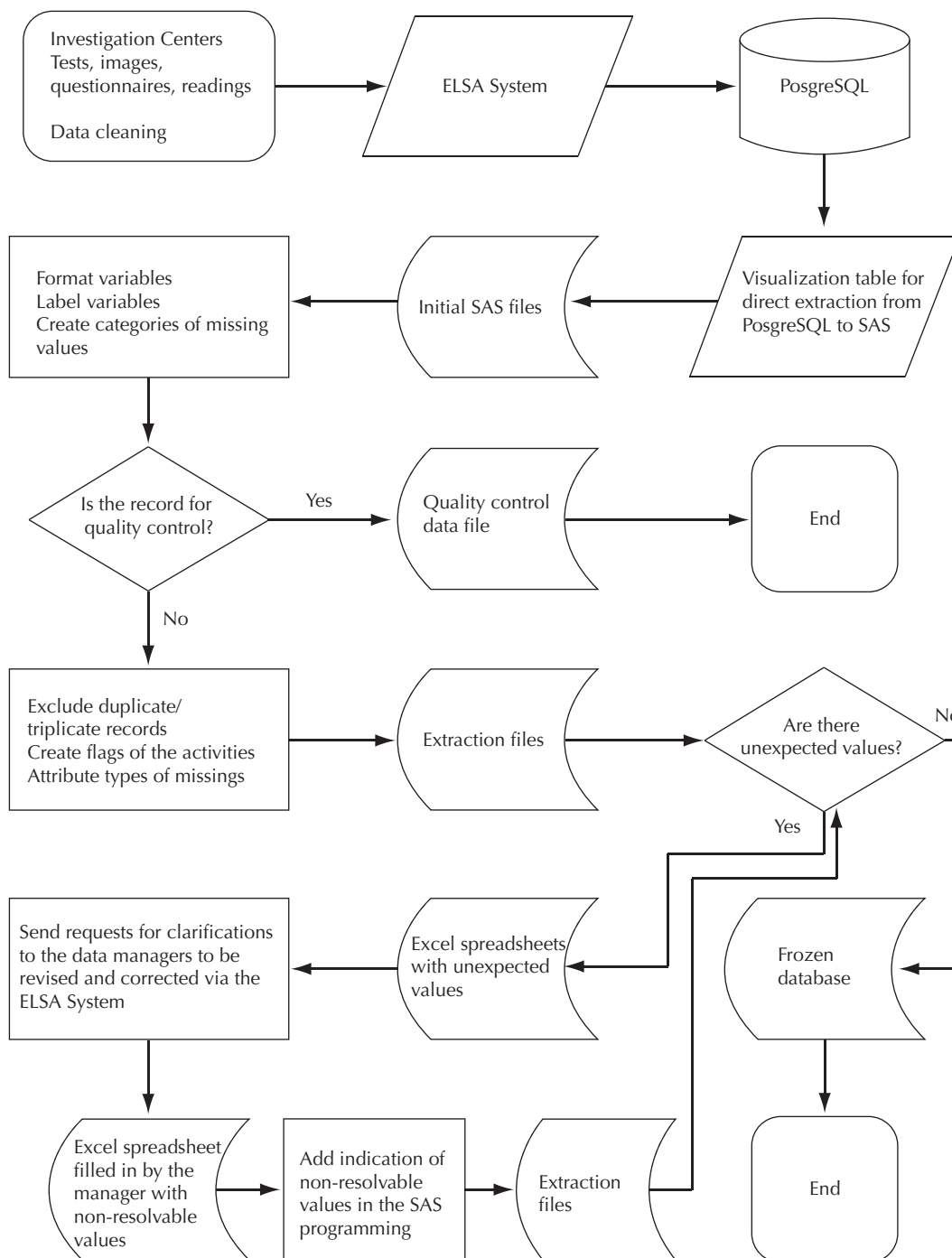


Figure 2. Action flow in data extraction and processing for the creation of statistical bases.

With the objective of preventing improper access to the system’s features, diverse security measures were taken:

- The web connections established between users and the system utilize an additional security layer (Secure Sockets Layer – SSL), transforming the HTTP connection into HTTPS. The issuer of the security certificate is the university itself, UFRGS, where the

server is located. Each use of the system generates a control of idle time. If the user spends too much time without interacting with the software, his session will expire, and he will be automatically disconnected. This type of timeout control prevents unauthorized access in cases in which the user leaves his computer unattended. In case his connection expires, he will have to identify himself again to access the system.

- After connecting, each user – besides having an identifier that is stored with the collection data – can only interact with the system through a pre-defined set of operations linked to different types of user profiles. The control of the access to specific features is performed by the creation and maintenance of these profiles.
- For the main interactions of users with the system, the information pertinent to the operation is additionally stored in text files (logs). Thus, it is possible to reverse some unsuccessful operations, and also to track different work flows and/or data flows executed within the system.

PACS-ELSA

The PACS-ELSA was developed to be an economic solution for the archiving of medical images, providing quick and decentralized access,³ and complying with the standard for manipulating, storing and transmitting medical information,^f called Digital Imaging and Communication in Medicine (DICOM). This standard includes the definition of the file format as well as the network communication protocol. The current version of DICOM was conceived in 1993 and has been internationally adopted by hospitals and clinics, creating proprietary and public domain solutions. Among the latter, the software dcm4chee^g was selected to develop the PACS-ELSA. This software, which applies the Health Level Seven (HL7) for image communication and archiving – the current communication standard of the ISO^h – was used to transmit retinography and echocardiography images. After the acquisition of images, they are sent to a central server to be accessed by the Reading Centers. Additionally, the results of the echocardiography reading are saved in the central server with the images by means of a DICOM object known as a Structured Report for transmission to the Datacenter and for future reviews and comparisons.

DATA SECURITY AND BACK-UP

The backup of the data of the ELSA System and of the PACS is carried out on a daily and weekly basis in an incremental modality and on a monthly basis in complete form, and it is stored on tape at a location outside of the CPD-UFRGS, together with the University's own backup datafiles.

SYSTEMS DOCUMENTATION

The Datacenter developed manuals to be used by the teams at the ICs, such as the Manual for Use of the System. For internal use, the entire development cycle was documented, using tools like *javadoc* and *wiki*. This documentation includes the programming stages, protocols such as implementation, and data extraction procedures.

OTHER COMPUTER TOOLS IN THE STUDY

The Datacenter also developed a SharePoint platform so that the members of the research team can share manuals and other documents. As presented in other papers of this supplement, the ELSA study also developed a website^b to disseminate the study to the public in general and to the participants, and a management system for the Publications Committee, through which proposals for papers are submitted for evaluation and approval.

CURRENT SITUATION AND PERSPECTIVES

Despite concern related to the use of a web system like the one developed, to our knowledge the first applied in Brazilian epidemiological research, only a minimal incidence of technical failures and inadequate utilization of the web instruments was observed, both with respect to direct entry of data online and to the keying of the paper forms. Additionally, with respect to direct, online entry during data collection (two centers), a low incidence of electrical failure and network problems was observed. As for keying, the biggest problem was a slow response time reported during some periods of the day. The experience acquired by the users and by the Datacenter in baseline data collection was fundamental for the consolidation of a system for longitudinal and multicenter use.

The ELSA system for data entry and management is probably the largest and most complex web system ever developed in Brazil for epidemiological research. Despite initial problems associated with delays in the availability of the system's features, the baseline data collected at the research centers for the 15,105 participants of the baseline are already in the ELSA System. The database in February 2012 contained 2,340 variables, in the SAS format, of 2,031,893 Kbytes. The Datacenter is still in the process of incorporating data from image readings, such as codes from readings of echocardiography and retinography images. Echocardiography images are available for approximately 10,000 participants and

^e Sociedade Brasileira de Informática em Saúde. Manual de requisitos de segurança, conteúdo e funcionalidades para sistemas de registro eletrônico em saúde (RES); versão 2.1. São Paulo: Conselho Federal de Medicina; 2004 [cited 2011 Mar 30]. Available from: http://www.uel.br/projetos/oicr/pages/arquivos/GTCERT_20040219_RT_V2.1.pdf

^f Medical Imaging & Technology Alliance. DICOM – Digital Imaging and Communications in Medicine. Rosslyn (VA); 2013 [cited 2013 Mar 4]. Available from: <http://medical.nema.org>

^g Evans D. dcm4chee. [cited 2013 Mar 4] Available from: <http://www.dcm4chee.org/confluence/display/ee2/Home>

^h ELSA-Brazil. [cited 2013 Mar 4] Available from: www.elsa.org.br

retinography images for more than 6000 in the PACS. The annual follow-up of the participants to monitor outcomes is being performed in the system at all the ICs. The acquired experience has enabled the enhancement of the web system for its effective utilization in future waves of data collection.

REFERENCES

1. Aquino EM, Barreto SM, Bensenor IM, Carvalho MS, Chor D, Duncan BB, et al. Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): objectives and design. *Am J Epidemiol*. 2012;175(4):315-24. DOI:10.1093/aje/kwr294
2. Berwanger O, presenter. Acetylcysteine for the prevention for contrast-induced nephropathy (ACT) Trial. Presented at the American Heart Association (AHA) Scientific Sessions; 2010 nov 16; Chicago (Ill).
3. Choplin RH, Boehme JM 2nd, Maynard CD. Picture archiving and communication systems: an overview. *Radiographics*. 1992;12(1):127-9.
4. Costa EA, Rose GA, Klein CH, Leal MC, Szwarcwald CL, Bassanesi SL, et al. Salt and blood pressure in Rio Grande do Sul, Brazil. *Bull Pan Am Health Organ*. 1990;24(2):159-76.
5. Duncan BB, Schmidt MI, Polanczyk CA, Homrich CS, Rosa RS, Achutti AC. Fatores de risco para doenças não-transmissíveis em área metropolitana na região sul do Brasil: prevalência e simultaneidade. *Rev Saude Publica*. 1993;27(1):43-8. DOI:10.1590/S0034-89101993000100007
6. Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring, and reporting. *Control Clin Trials*. 1995;16(2 Suppl):104S-36S.
7. Guimarães R, Lourenço R, Cosac S. A pesquisa em epidemiologia no Brasil. *Rev Saude Publica*. 2001;35(4):321-40. DOI:10.1590/S0034-89102001000400001
8. Kent B, Andres C. Extreme Programming explained: embrace change. Boston (MA): Addison-Wesley; 2000.
9. Malerbi DA, Franco LJ. Multicenter study of the prevalence of diabetes mellitus and impaired glucose tolerance in the urban Brazilian population aged 30-69 yr. The Brazilian Cooperative Group on the Study of Diabetes Prevalence. *Diabetes Care*. 1992;15(11):1509-16.
10. Schmidt JR, Vignati AJ, Pogash RM, Simmons VA, Evans RL. Web-based distributed data management in the childhood asthma research and education (CARE) network. *Clin Trials*. 2005;1;2(1):50-60. DOI:10.1191/1740774505cn630a
11. Schmidt MI, Duncan BB, Reichelt AJ, Branchtein L, Matos MC, Costa e Forti A, et al. Gestational diabetes mellitus diagnosed with a 2-h 75-g oral glucose tolerance test and adverse pregnancy outcomes. *Diabetes Care*. 2001;24(7):1151-5. DOI:10.2337/diacare.24.7.1151
12. Schwaber K. Agile project management with scrum. Redmond (WA): Microsoft Press; 2004.
13. Winget M, Kincaid H, Lin P, Li L, Kelly S, Thornquist M. A web-based system for managing and coordinating multiple multisite studies. *Clinical Trials*. 2005;2(1):42-9. DOI:10.1191/1740774505cn620a

The Brazilian Longitudinal Study for Adult Health (ELSA-Brasil) was funded by the Ministry of Health (DECIT – Departamento de Ciência e Tecnologia – Science and Technology Department) and by the Ministry of Science and Technology (FINEP – Financiadora de Estudos e Projetos [Funding Agency for Studies and Projects] and CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development]) (Processes N. 01 06 0010.00 RS, 01 06 0212.00 BA, 01 06 0300.00 ES, 01 06 0278.00 MG, 01 06 0115.00 SP, 01 06 0071.00 RJ).

The authors declare that there are no conflicts of interest.

This manuscript was submitted for publication and underwent a peer review process as any other manuscripts submitted to this publication, and anonymity was guaranteed for authors and reviewers. Editors and reviewers declare no conflicts of interest that may affect the peer-review process.

ACKNOWLEDGMENT

To Dr. Lloyd Chambless (University of North Carolina at Chapel Hill) for his support and guidance in the establishment of the ELSA Datacenter over the past years.