



Revista de Saúde Pública

ISSN: 0034-8910

revsp@usp.br

Universidade de São Paulo

Brasil

Duncan, Bruce Bartholow; Vigo, Álvaro; Hernandez, Émerson; Luft, Vivian Cristine; Ahlert, Hubert; Bergmann, Kaiser; Mota, Eduardo

Gerência de Informação em Estudos Multicêntricos: o Estudo Longitudinal de Saúde do Adulto

Revista de Saúde Pública, vol. 47, núm. 2, junio, 2013, pp. 95-104

Universidade de São Paulo

São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=67240211013>

- ▶ [Como citar este artigo](#)
- ▶ [Número completo](#)
- ▶ [Mais artigos](#)
- ▶ [Home da revista no Redalyc](#)

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Bruce Bartholow Duncan<sup>I</sup>Álvaro Vigo<sup>I</sup>Emerson Hernandez<sup>I</sup>Vivian Cristine Luft<sup>I</sup>Hubert Ahlert<sup>II</sup>Kaiser Bergmann<sup>III</sup>Eduardo Mota<sup>IV</sup>

# Gerência de Informação em Estudos Multicêntricos: o Estudo Longitudinal de Saúde do Adulto

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

### RESUMO

A gerência da informação em estudos multicêntricos de grande porte requer uma abordagem especializada. O Estudo Longitudinal da Saúde do Adulto (ELSA-Brasil) criou um Centro de Dados para delinear e gerenciar seu sistema de dados. O objetivo do artigo foi descrever os passos envolvidos, incluindo os métodos de entrada, transmissão e gerência de informações. Foi desenvolvido um sistema *web* que permitiu, de forma segura e confidencial, a entrada *online*, verificação e edição, bem como incorporação de dados coletados em papel. Além disso, foi implantado e personalizado um sistema de armazenamento e comunicação de imagens (*Picture Archiving and Communication System*) para ecocardiografia e retinografia que armazena as imagens recebidas dos Centros de Investigação e as torna acessíveis nos Centros de Leitura. Finalmente, foram desenvolvidos processos de extração e limpeza de dados para criação de bases de dados em formatos que permitam análises em múltiplos pacotes estatísticos.

<sup>I</sup> Programa de Pós-Graduação em Epidemiologia. Faculdade de Medicina. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brasil

<sup>II</sup> Centro de Processamento de Dados. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brasil

<sup>III</sup> Faculdade da Medicina. Universidade Federal de Minas Gerais. Belo Horizonte, MG, Brasil

<sup>IV</sup> Instituto de Saúde Coletiva. Universidade Federal da Bahia. Salvador, BA, Brasil

**DESCRITORES:** Estudos Multicêntricos como Assunto, métodos. Bases de Dados como Assunto. Sistemas de Gerenciamento de Base de Dados. Sistemas de Informação. Armazenamento e Recuperação da Informação. Estudos de Coortes.

### Correspondência | Correspondence:

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

Recebido: 5/10/2011  
Aprovado: 5/6/2012

Artigo disponível em português e inglês em:  
[www.scielo.br/rsp](http://www.scielo.br/rsp)

---

## 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.

---

## INTRODUÇÃO

Em 2001, Guimarães et al<sup>7</sup> descreveram que algumas áreas da produção científica da epidemiologia brasileira eram ainda pouco expressivas, como a das doenças crônicas não transmissíveis. Na última década, em resposta a um aumento substancial no fomento à pesquisa em saúde, essas doenças passaram a ser mais e mais bem investigadas. A abrangência das informações colhidas foi ampliada e a metodologia para sua aquisição cresceu em complexidade.

Um exemplo do novo cenário da pesquisa epidemiológica brasileira é a criação de rede de pesquisadores para construção do Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil),<sup>1</sup> estudo de coorte que objetiva acompanhar 15.000 adultos ao longo do tempo em seis centros brasileiros. As avaliações envolvem entrevistas e exames de complexidades variáveis, incluindo imagens. Aspectos como o número de participantes e de formulários de entrevistas e exames, natureza longitudinal, organização multicêntrica e avaliações complexas do ELSA foram determinantes na definição das características de seu sistema de dados.

No final da década de 1970, um dos primeiros estudos epidemiológicos em doenças crônicas realizados no País incluiu adultos em uma amostra probabilística do Rio Grande do Sul. As informações foram obtidas por meio de entrevistas domiciliares e da medida da pressão arterial, e foram anotadas em papel para posterior digitação e entrada em computador de grande porte (*mainframe*), típico da época.<sup>4</sup> Metodologia semelhante foi empregada na década

seguinte, no Estudo da Prevalência do Diabetes, um estudo multicêntrico, em vários estados brasileiros, com digitação de dados centralizada em São Paulo.<sup>9,a</sup> Outro estudo naquela década, o Estudo de Fatores de Risco para Doenças Crônicas Não Transmissíveis, registrou suas informações em papel, e a digitação foi feita em microcomputadores.<sup>5</sup> Na década de 1990, um estudo multicêntrico sobre diabetes na gravidez utilizou EpiInfo para verificação de inconsistências na digitação dupla, para controle de erros de pulos nas entrevistas e para detecção de valores incongruentes nas medidas.<sup>11</sup>

O registro em papel das informações nesses estudos requeria revisão e codificação por supervisor do projeto antes da digitação, o que muitas vezes ocorria tarde, quando não era mais possível corrigir os erros detectados. Para reduzir esse problema, o Estudo de Consumo e Comportamento Alimentar na Gestação utilizou um sistema que eliminava a etapa da digitação, controlava erros de pulos nas entrevistas e detectava valores incongruentes. O método era escanear os questionários e formulários para captação dos dados pelo programa Teleform (Cardiff, Vista CA, EUA). O procedimento não eliminava a necessidade de se utilizar papel para o registro das informações e, quando o escaneamento era retardado, o potencial de controle de erros ficava limitado. Alguns ensaios clínicos randomizados, cujo volume de informação coletada de cada indivíduo é em geral pequeno, passaram a utilizar sistemas *web* para entrada e gerência de dados.<sup>2</sup>

---

<sup>a</sup> Centers for Disease Control and Prevention. EpiInfo Version 6: a word processing, database and statistics program for public health on IBM-compatible microcomputers [programa de computador]. Atlanta; 1995.

A Tabela 1 destaca a evolução dos procedimentos de coleta e entrada de dados nesses estudos e em alguns estudos epidemiológicos clássicos internacionais sobre doenças crônicas no mesmo período. Observa-se a tendência para o uso de sistemas com entrada de dados em microcomputadores, por garantirem melhor confidencialidade dos dados e por permitirem a detecção e correção imediata de erros, muitas vezes, enquanto o indivíduo ainda está com a equipe de campo.<sup>6</sup>

Quando a coleta dos dados é feita em ambiente de Centro de Pesquisa, a tendência é utilizar entrada via *web* conectada a um sistema de gerência da informação.<sup>10,13</sup> Além de possibilitar maior confiabilidade dos dados, esses sistemas podem alertar para a presença de situações de interesse (por exemplo, diabetes) e direcionar ações específicas de acordo com características dos indivíduos da pesquisa (por exemplo, fez cirurgia bariátrica, não faz teste de tolerância à glicose). Com monitoramento periódico e efetivo dos dados e pronta correção, esses sistemas permitem também que os dados sejam rapidamente transformados em bases de análise.

O sistema desenhado para o ELSA procurou incorporar as vantagens de sistemas *web*. O objetivo deste artigo foi mostrar a estruturação do Centro de Dados do ELSA e descrever os métodos de entrada, transmissão e gerência de informações no estudo.

## CENTRO DE DADOS

O Centro de Dados no ELSA é composto por uma equipe multidisciplinar das áreas da epidemiologia, estatística, tecnologia da informação (analistas, programadores e *webdesigner*) e relações públicas. Um dos objetivos do projeto é a formação de recursos humanos para pesquisa. O Centro de Dados oportuniza a participação de bolsistas de graduação e de pós-graduação, agrupando-os em três equipes distintas (Figura 1): Bioestatística e Epidemiologia, Programação Estatística e Sistemas. O Centro conta com Consultores do Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul (CPD-UFRGS) para suporte técnico e peritos internacionais (membros da equipe do *Collaborative Studies Coordinating Center* da *University of North Carolina*).

**Tabela 1.** Sistemas de coleta e entrada de dados em alguns estudos epidemiológicos brasileiros e internacionais sobre doenças crônicas.

Estudo	Ano	n	Sistema de coleta e entrada de dados
<b>Brasileiros</b>			
<i>Salt and Blood Pressure in Rio Grande do Sul, Brasil</i>	1978	4.565	Papel + digitação mainframe
Estudo de Fatores de Risco para Doenças Crônicas Não Transmissíveis – RS	1986	1.157	Papel + digitação microcomputador
Estudo da Prevalência do Diabetes	1986	21.847	Papel + digitação mainframe
Projeto Bambuí	1996	1.606	Papel + digitação microcomputador
Estudo Pró-Saúde	1999	4.030	Papel + digitação microcomputador
ECCAGE	2006	2.590	Papel + escaneamento (Teleform)
ACT	2008	2.308	Sistema web
<b>Internacionais</b>			
ARIC (Estados Unidos)	1987	15.792	Entrada direta microcomputadores; posteriormente, sistema web
CHS (Estados Unidos)	1989	5.201	Entrada direta microcomputadores; posteriormente, sistema web
AUSDIAB (Austrália)	1999	1.478	Papel + escaneamento (Teleform)
MESA (Estados Unidos)	2000	6.814	Papel + escaneamento (Teleform); posteriormente, sistema web
Kadoorie (China)	2004	500.000	Entrada direta microcomputadores
BioBank (Reino Unido)	2007	500.000	Entrada direta microcomputadores
HCHS/SOL (Estados Unidos)	2008	16.000	Sistema web

ECCAGE: Estudo do Consumo e Comportamento Alimentar na Gestação

ACT: Acetylcysteine for the Prevention of Contrast-Induced Nephropathy

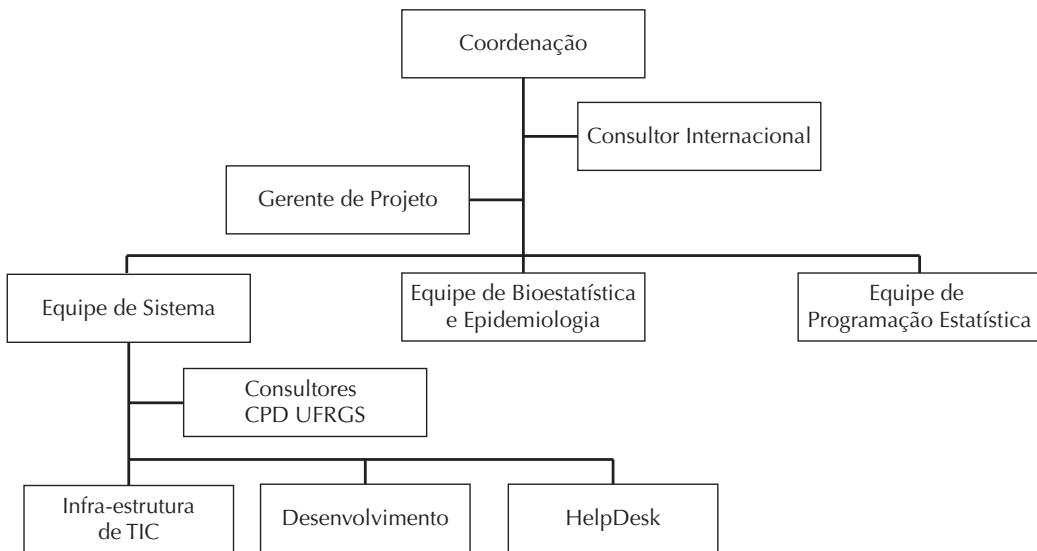
ARIC: Atherosclerosis Risk in Communities Study

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



TIC: Tecnologia de Informação e Comunicação

**Figura 1.** Estrutura do Centro de Dados do ELSA (CPD-UFRGS: Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul).

Para acompanhar as atividades do Centro de Dados, foi criado um Núcleo de Desenvolvimento do Sistema, composto por representantes dos Centros de Investigação (CIs) com experiência em tecnologia da informação e desenvolvimento de sistemas e por dois membros do próprio Centro de Dados.

O coordenador do Centro de Dados é membro do Comitê Diretivo do projeto e participa do planejamento geral do estudo. As atividades do Centro de Dados na linha de base foram focadas na criação de mecanismos para a entrada, processamento e limpeza dos dados, no recebimento de imagens dos Centros de Investigação e seu encaminhamento aos Centros de Leitura, e na criação e distribuição das bases de dados do estudo. O elo operacional entre os CIs e o Centro de Dados é feito por gerentes de dados.

## OS SISTEMAS DE INFORMAÇÕES DO ESTUDO

Os sistemas de informação desenvolvidos consistem de módulos *web* – o Sistema ELSA –, que disponibilizam os formulários de aquisição, e de um sistema de comunicação e arquivamento de imagens (em inglês, *Picture Archiving and Communication System*, Pacs) – o Pacs-ELSA.

Uma das primeiras decisões tomadas foi pela utilização de *software* livre. Vantagens quanto a preço, modelo de negócios e suporte são apenas alguns dos fatores que pesaram nessa decisão. Destaca-se também que o ELSA

seguiu a orientação do governo federal de desenvolver *software* sob seu domínio de forma livre e aberta.<sup>b</sup>

Os sistemas estão hospedados fisicamente no CPD-UFRGS, onde foram criadas máquinas virtuais específicas para as necessidades de desenvolvimento, teste e produção. Por medidas de segurança, foi criada uma máquina específica para hospedar o banco de dados de produção, acessível apenas via computadores da rede UFRGS. A comunicação com os demais centros ocorre via Rede Nacional de Pesquisa (RNP),<sup>c</sup> permitindo boa velocidade de transmissão, largura de banda e disponibilidade. O Pacs funciona na mesma máquina virtual de produção.

## O SISTEMA ELSA

### Opções de Desenvolvimento e Características do Sistema ELSA

Uma das primeiras escolhas no desenvolvimento do Sistema ELSA foi utilizar *software* livre, descartando o uso de ferramentas e tecnologias pagas, como o C# da Microsoft®. Outra decisão inicial, considerando o caráter longitudinal da pesquisa e a diversidade na origem dos dados, foi não utilizar *softwares desktop/offline*, como Epi6, nem sistemas cliente-servidor (por exemplo, como o utilizado para a declaração de imposto de renda em anos recentes), bem como sistemas de captação de imagens de formulários em papel (como o Cardiff Teleform) e outras modalidades listadas na Tabela 1.

<sup>b</sup> Brasil. Governo Federal. Diretrizes de uso de Software Livre no Governo Federal. Brasília (DF); 2011 [citado 2011 mar 30]. Disponível em: <http://www.softwarelivre.gov.br/planejamento-cisl/diretrizes>

<sup>c</sup> Rede Nacional de Pesquisa. Relatório de gestão RNP: edição anual 2012. Rio de Janeiro; 2012 [citado 2013 mar 04]. Disponível em: [http://www.rnp.br/rnp/relatorio\\_gestao.html](http://www.rnp.br/rnp/relatorio_gestao.html) 201

A adoção do sistema *web* permitiu acesso automático não apenas às regras de validação na entrada de dados, mas também às informações já armazenadas no sistema (por exemplo, formas preferenciais de contato com os participantes). Como um sistema *web* com as funcionalidades requeridas não estava disponível em domínio público, optou-se pela criação de um sistema novo com apoio do CPD-UFRGS. A plataforma Java foi escolhida como ferramenta de desenvolvimento e ambiente de execução pela flexibilidade oferecida. A Tabela 2 apresenta uma lista com as principais ferramentas livres usadas.

No Sistema ELSA, o usuário pode utilizar abas e menus para acessar os formulários de pesquisa, tais como para recrutamento, recepção, questionários e exames. Há também uma interface de digitação de dados coletados inicialmente em papel, incluindo dados de auditoria (data, hora e técnico responsável pela coleta). Na entrada direta *online* essas informações são captadas automaticamente.

Para os exames e entrevista na linha de base (Onda 1), foram criados 38 formulários, todos disponibilizados *online* e em papel para preenchimento na presença de problemas técnicos (locais de coleta fora da sede, momentos de falha na rede elétrica ou da rede para transmissão ao servidor central). Aspectos do delineamento do estudo, população estudada, recrutamento e exames e entrevistas realizadas na linha de base foram detalhados em outra publicação.<sup>1</sup>

Por ser um sistema integrado com compartilhamento de informações em todos os locais de entrada de dados, o Sistema ELSA oferece ferramentas de gerência que auxiliam na comunicação entre os centros e na administração da coleta de dados nos CIs (por exemplo, dados de alteração de cadastro e os obtidos na aquisição de exames especializados são disponibilizados automaticamente para os centros de leitura).

## Processo de Desenvolvimento do Sistema

O processo de desenvolvimento do sistema ELSA foi incremental, tendo como núcleo as etapas de análise, programação e teste. Nessa estratégia, chamada de desenvolvimento interativo e incremental, a construção de *software* inicia com um protótipo pequeno e, mediante sucessivos refinamentos, é construído o sistema total. A estratégia foi muito útil no caso do ELSA, uma vez que a novidade do processo não permitiu que os requisitos do sistema estivessem todos elucidados em um primeiro momento. Além disso, possibilitou que todos os envolvidos no processo de programação descobrissem detalhes importantes em estágios menos avançados do desenvolvimento, tornando mais fácil sua mudança e/ou adaptação.<sup>4</sup> Além da estratégia incremental, o desenvolvimento também utilizou métodos ágeis, como *Scrum* para gestão e planejamento de projetos<sup>12</sup> e *XP (Extreme Programming)* para desenvolvimento de *software*.<sup>8</sup> A programação foi feita em iterações, em geral de três ou quatro semanas.

À medida que o sistema foi sendo desenvolvido, constatou-se a importância de contar com um gerente de produto responsável pelo desenvolvimento de uma “visão do produto”, o que permitiu realizar maior discussão inicial sobre as necessidades gerais do sistema, antes de entrar na programação de suas partes. A visão do produto foi importante para identificar interessados, dentro e fora do Centro de Dados, e delinear as necessidades previstas. A partir disso, a Equipe de Sistemas traduziu as necessidades em uma proposta de funcionalidades para o sistema. Após revisão e aprovação de cada atividade pelo gerente e pelo pesquisador responsável, o Centro de Dados iniciou a programação do instrumento específico.

## Montagem dos Formulários

Os principais passos do processo de montagem de formulário de coleta no sistema estão summarizados

**Tabela 2.** Principais ferramentas livres utilizadas para o desenvolvimento do sistema ELSA.

Elemento	Tecnologia	Disponível em
Servidor de Aplicação	JBoss	<a href="http://www.jboss.org/">www.jboss.org/</a>
Sistema Operacional	Debian	<a href="http://www.debian.org">www.debian.org</a>
Criação de Questionários	Opinio	<a href="http://objectplanet.com/opinio/">objectplanet.com/opinio/</a>
Ambiente de Desenvolvimento	Eclipse	<a href="http://www.eclipse.org/">www.eclipse.org/</a>
SGBD	PostgreSQL	<a href="http://www.postgresql.org/">www.postgresql.org/</a>
Linguagem de Programação	Java	<a href="http://www.java.com/">www.java.com/</a>

SGBD: Sistema de Gerenciamento de Banco de Dados

<sup>4</sup> Beck K, Beedle M, Bennekum A, Cockburn A, Cunningham W, Fowler M, et al. Manifesto para o desenvolvimento ágil de software.[citado 2011 mar 2]. Disponível em: <http://manifestoagil.com.br/>

na Tabela 3. A montagem inicial desses formulários foi feita em papel pelos pesquisadores ELSA, responsáveis pelos instrumentos correspondentes. Equipe do Centro de Dados revisou e padronizou aspectos como regras de pulo, tratamento de valores faltantes e limites mínimo e máximo das variáveis, registrando o processo em um documento chamado de mapas de variáveis. As regras de obrigatoriedade de preenchimento estabelecidas foram, em geral, as seguintes: para os questionários, optou-se por preenchimento obrigatório e por valores mínimo e máximo fixos; para formulários de exames, optou-se por sinalizar a presença de valores fora dos limites mínimo e máximo estabelecidos e de dados faltantes, e por exigir sua confirmação antes de prosseguir. Uma vez estruturado, padronizado e aprovado, o formulário foi encaminhado à Equipe de Sistemas para traduzir as especificações para uma representação no universo de *software* e programar as funcionalidades previstas.

A próxima etapa foi a realização de testes internos das funcionalidades produzidas, o que foi feito pela Equipe de Sistemas (testes de integração e exploratórios) e pela Equipe de Bioestatística e Epidemiologia (testes de aceitação em uma máquina de testes). Os testes de integração e exploratórios eram feitos logo após a programação. Uma vez considerados prontos,

**Tabela 3.** Etapas da disponibilização de funcionalidades no sistema de dados.

- A Equipe de Bioestatística e Epidemiologia entra em contato com o pesquisador responsável por cada instrumento (formulário, questionário, relatório, e outros) para colher as especificações necessárias para sua implementação no sistema.
- A Equipe de Programação Estatística revisa e padroniza o formulário, especificamente aspectos como regras de pulo e tratamento de valores faltantes e limites, gerando mapas de variáveis.
- O coordenador do Centro de Dados estabelece prioridades de implementação de novas funcionalidades no sistema (com base nas demandas do Comitê Diretivo).
- A Equipe de Sistemas traduz as especificações contidas nos mapas para uma representação no universo de software, e programa, faz testes iniciais e disponibiliza a nova funcionalidade para testes pela Equipe de Bioestatística e Epidemiologia.
- Após finalizar os testes, a Equipe de Bioestatística e Epidemiologia entra novamente em contato com o pesquisador ELSA responsável pelo instrumento para verificar a aceitabilidade e homologar a funcionalidade no sistema.
- A Equipe de Sistemas disponibiliza no servidor de produção as funcionalidades para entrada de dados e também as funcionalidades que permitem sua extração.
- A Equipe de Programação Estatística, seguindo orientação da Equipe de Bioestatística e Epidemiologia, extrai, processa, gerencia a limpeza dos dados extraídos e elabora dicionários de dados e outros documentos para viabilizar a análise estatística.

os itens eram disponibilizados para os testes de aceitação. Esse segundo grupo de testes verificou, a cada final de ciclo, se o que foi implementado estava de acordo com as necessidades do projeto. Os testes foram gravados e salvos para uso posterior pela Equipe de Sistemas, evitando a necessidade de a Equipe de Bioestatística e Epidemiologia repetir os mesmos testes no futuro.

Uma vez aprovada internamente, a funcionalidade era avaliada e homologada pelo pesquisador responsável para entrada em uma implantação na máquina de produção, integrada à parte do sistema e pronta para uso.

### Uso do Sistema

Após o desenvolvimento e homologação dos instrumentos de coleta de dados, os membros da equipe ELSA nos diversos CIs alimentavam o sistema de forma *online* ou a partir de uma digitação do formulário correspondente em papel. Um sistema de apoio, tipo HelpDesk aos usuários, foi criado para esclarecer dúvidas, receber relatos de potenciais *bugs* de programação e responder aos encaminhamentos efetuados. O HelpDesk é disponibilizado por telefone e por e-mail específico, sendo exercido por estagiário de relações públicas.

### Extração e Preparação de Bases para Análise

As informações coletadas via Sistema ELSA são armazenadas em um banco de dados PostgreSQL®, com tabelas de dados separadas para cada atividade. A extração e transferência desses dados para bases de análise estatística foram realizadas periodicamente, em geral, a cada mês. O processamento da extração de dados é realizado com o programa SAS® (*Statistical Analysis System*), envolvendo várias etapas (Figura 2). Os dados de cada atividade são extraídos individualmente, utilizando um apontamento direto entre o SAS e uma tabela de visualização dos dados, produzida pela Equipe de Sistemas a partir do PostgreSQL®. As extrações são organizadas por data e os dados de cada atividade são armazenados em pasta correspondente. Ao longo da entrada de dados, foram elaborados relatórios para informar os pesquisadores ELSA sobre o andamento do processo.

### Monitoramento da Entrada de Dados

Atividades de monitoramento e limpeza foram acopladas ao processo de extração.<sup>6</sup> Para acompanhar a entrada de dados de cada participante, foi criada uma variável indicadora da realização de cada atividade, chamada de *flag*. Quando não existem dados da atividade no sistema, a variável fica em branco, o que permite identificar quais participantes ainda não têm tal atividade preenchida. Quando sabidamente um participante não realizou uma

ou mais atividades, um valor especial pode ser atribuído a essa variável indicadora.

Quando unidas, as variáveis indicadoras (*flags*) formam um arquivo chamado de “Relatório de *Flags*”, que é armazenado em formato SAS (sas7bdat) e .xls. O Relatório de *Flags* no formato .xls é encaminhado aos CIs para identificação de participantes com atividades incompletas.

## Limpeza de Dados

Na extração, são atribuídos valores especiais para identificar tipos específicos de não resposta, tais como pulos quando a questão ou exame não se aplica; recusa em responder alguma questão e outros. Dessa forma, valores não esperados, tais como valores faltantes (*missings*) ou inconsistentes, podem ser identificados

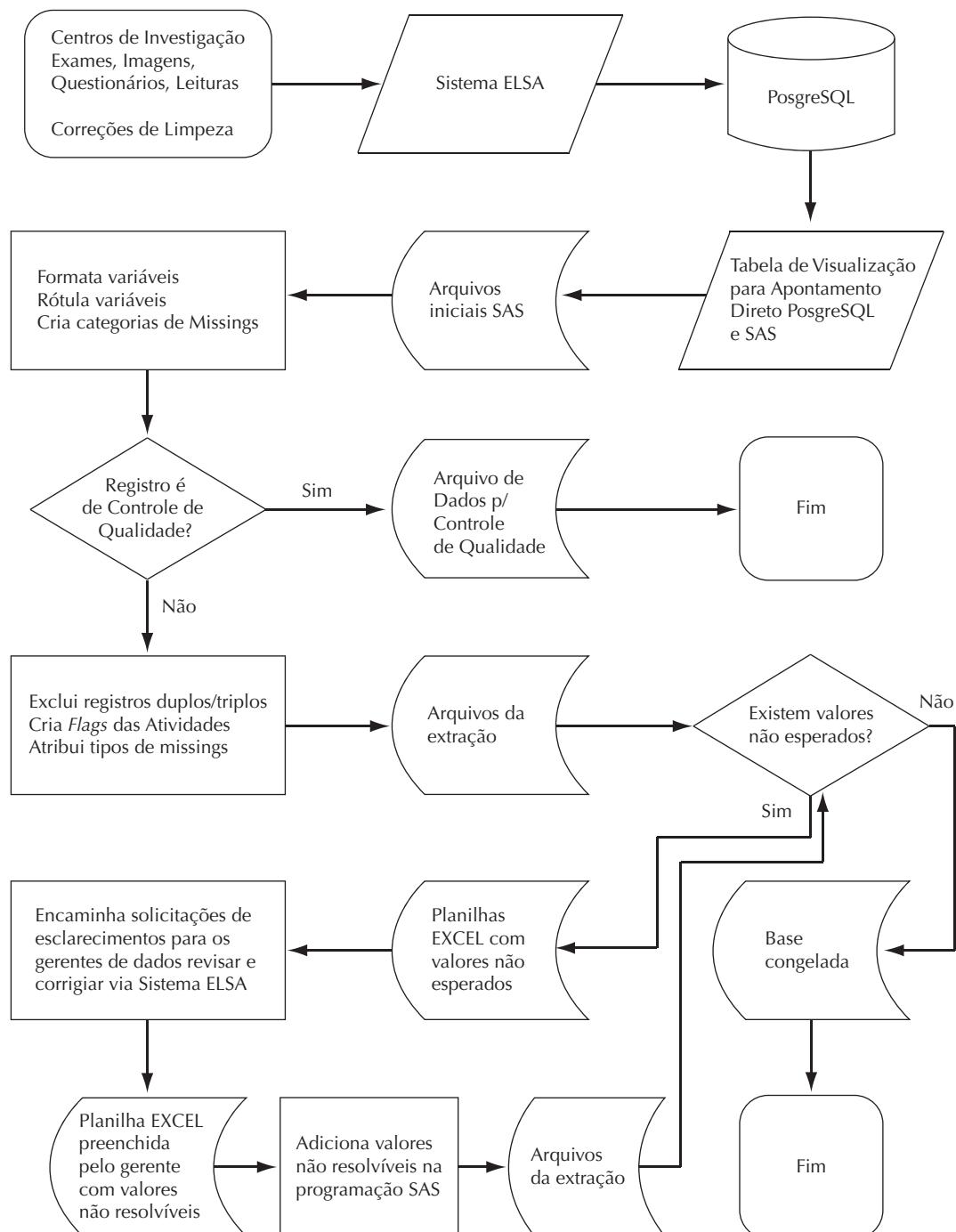


Figura 2. Fluxo de ações na extração e processamento dos dados para criação das bases estatísticas.

e comunicados aos CIs para elucidação. Um relatório de inconsistências potenciais é criado após a extração e enviado para revisão nos CIs. O gerente de dados do CI comunica os resultados dessa revisão ao Centro de Dados e efetua as correções possíveis, utilizando o próprio sistema *web* de entrada de dados.

### Geração de Bases para Análise

Quando todas as informações estiverem inseridas no sistema e as etapas de monitoramento e limpeza estiverem concluídas, a base de dados em SAS será “congelada” para dar início às análises formais do estudo. Bases preliminares foram distribuídas aos coordenadores dos CIs, periodicamente, para permitir exame preliminar dos dados. Nessas bases, os participantes são rotulados com números identificadores diferentes daqueles utilizados no processo de coleta. Os arquivos SAS gerados na extração são gravados também em outros formatos, como, por exemplo, SAV (SPSS) e DTA (Stata), permitindo que os pesquisadores efetuem análises em SAS, SPSS, Stata, R e outros programas.

Também foi criado um “livro dos dados” para cada atividade, contendo uma descrição sucinta das informações de cada variável. Para as variáveis quantitativas são apresentados o número de observações válidas, os valores mínimo e máximo, a média e o desvio padrão. Para as variáveis qualitativas, são apresentadas as frequências de cada categoria, o número de observações válidas e de valores faltantes. Um sistema com interface gráfica possibilitará ao pesquisador selecionar as variáveis desejadas para a criação de um banco específico de dados para análise, no formato de arquivo desejado. O sistema poderá gerar o banco de dados e o dicionário correspondente de variáveis, para gravação em local previamente especificado.

### Aspectos de Segurança no Uso do Sistema

A Sociedade Brasileira de Informática em Saúde<sup>e</sup> apresenta um conjunto de requisitos, classificados como obrigatórios e desejáveis, adaptados da International Standards Organization (ISO), que definem algumas características básicas para a construção de um bom sistema de registro eletrônico em saúde. O ELSA implementou o sistema de registro eletrônico em saúde, contemplando, na medida do possível, os requisitos obrigatórios e os desejáveis, como descrito a seguir. Com o objetivo de prevenir acesso indevido às funcionalidades do sistema, diversas medidas distintas de segurança foram tomadas:

- As conexões web estabelecidas entre os usuários e o sistema usam uma camada de segurança adicional (*Secure Sockets Layer – SSL*), transformando a conexão HTTP em HTTPS. O emissor do certificado de segurança é a própria universidade, UFRGS, onde o servidor está localizado. Cada uso do sistema gera um controle de tempo de ociosidade. Caso o usuário passe muito tempo sem interagir com o *software* ele será automaticamente desconectado, pois sua sessão terá expirado. Esse tipo de controle por *timeout* acontece para prevenir acesso não autorizado em casos em que o usuário deixa seu computador sem supervisão. Caso sua conexão termine, ele deverá novamente identificar-se no sistema.
- Após conectar-se, cada usuário – além de ter um identificador que é armazenado com os dados de coleta – só pode interagir com o sistema por um conjunto pré-definido de operações vinculadas a diferentes tipos de perfis de usuários. O controle de acesso às funcionalidades é feito pela criação e manutenção desses perfis.
- Para as principais interações dos usuários com o sistema, as informações pertinentes à operação são armazenadas adicionalmente em arquivos texto (*logs*). Com isso, é possível reverter alguma operação malsucedida, assim como rastrear diferentes fluxos de trabalho e/ou dados executados dentro do sistema.

### Pacs-ELSA

O Pacs-ELSA foi desenvolvido para ser uma solução econômica de armazenamento de imagens médicas, provendo acesso rápido e descentralizado,<sup>3</sup> dentro do padrão para manipulação, armazenamento e transmissão de informação médica,<sup>f</sup> denominado *Digital Imaging and Communication in Medicine* (Dicom). Esse padrão inclui a definição do formato de arquivo, assim como o protocolo de comunicação de rede. A versão atual do Dicom foi concebida em 1993 e vem sendo adotada internacionalmente por hospitais e clínicas, criando soluções proprietárias e livres. Dentre as soluções livres e abertas existe a ferramenta dcm4chee,<sup>g</sup> usada para desenvolver o Pacs-ELSA. Esse sistema, que aplica o padrão *Health Level Seven* (HL7) para comunicação e armazenamento das imagens, atual padrão de comunicação do ISO,<sup>f</sup> foi utilizado para transmissão de imagens de retina e de ecocardiografia. Após a aquisição de imagens, elas são encaminhadas a um servidor central para serem acessadas pelos Centros de Leitura. Além disso, os resultados da leitura de ecocardiografia são gravados

<sup>e</sup> 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 [citado 2011 mar 30]. Disponível em: [http://www.uel.br/projetos/oicr/pages/arquivos/GTCERT\\_20040219\\_RT\\_V2.1.pdf](http://www.uel.br/projetos/oicr/pages/arquivos/GTCERT_20040219_RT_V2.1.pdf)

<sup>f</sup> Medical Imaging & Technology Alliance. DICOM – Digital Imaging and Communications in Medicine. Rosslyn (VA); 2013 [citado 2013 mar 4]. Disponível em: <http://medical.nema.org>

<sup>g</sup> Evans D. dcm4chee. [citado 2013 mar 4]. Disponível em: <http://www.dcm4chee.org/confluence/display/ee2/Home>

no servidor central, com as imagens, por meio de objetos Dicom conhecidos como *Structured Reports* para transmissão ao Centro de Dados e para revisões e comparações futuras.

### **SEGURANÇA E BACK-UP DOS DADOS**

Cópias de segurança (*backup*) dos dados do Sistema ELSA e do Pacs são realizadas diária e semanalmente na modalidade incremental, e mensalmente na modalidade completa, e armazenadas em fita em um local fora do CPD-UFRGS, juntamente com cópias dos dados da própria universidade.

### **DOCUMENTAÇÃO DOS SISTEMAS**

O Centro de Dados desenvolveu manuais para uso das equipes nos CIs, como, por exemplo, o Manual de Uso do Sistema. Para uso interno, foi documentado todo o ciclo de desenvolvimento, utilizando ferramentas como *javadoc* e *wiki*. Essa documentação inclui desde etapas de programação a protocolos, como implantação e procedimentos de extração dos dados.

### **OUTRAS FERRAMENTAS DE INFORMÁTICA NO ESTUDO**

O Centro de Dados desenvolveu também uma plataforma SharePoint para compartilhar manuais e outros documentos aos membros da equipe executora. Como apresentado em outros artigos deste suplemento, o estudo ELSA desenvolveu também um *website*<sup>h</sup> de divulgação do estudo ao público em geral e aos participantes, e um sistema de gerência para o Comitê de Publicações, ao qual as propostas de artigos são submetidas para avaliação e aprovação.

### **SITUAÇÃO ATUAL E PERSPECTIVAS**

Apesar do receio na utilização de um sistema *web* como esse, inédito em pesquisa epidemiológica

brasileira, cabe destacar que, tanto para a entrada direta de dados *online* quanto para a digitação dos formulários em papel, se observou incidência mínima de falhas técnicas ou utilização inadequada dos instrumentos *web* após sua homologação. Para a entrada direta *online* durante a coleta de dados (dois centros), observou-se baixa incidência de falta de energia elétrica e problemas de rede. Para a digitação, o problema maior encontrado foi a lentidão relatada em alguns períodos do dia. A experiência adquirida pelos usuários e pelo Centro de Dados na linha de base foi fundamental para a consolidação de um sistema para uso longitudinal e multicêntrico.

O sistema de entrada e gerência de dados ELSA-Brasil é provavelmente o maior e mais complexo sistema *web* desenvolvido no Brasil para pesquisa epidemiológica. Apesar de problemas iniciais associados aos atrasos na disponibilização de funcionalidades do sistema, os dados obtidos nos CIs sobre os 15.105 participantes da linha de base já estão no Sistema ELSA. A base de dados em fevereiro de 2012 continha 2.340 variáveis; sendo, no formato SAS, de 2.031.893 *Kbytes*. Ainda falta incorporar vários resultados de leitura de imagens, por exemplo, da ecocardiografia e da retina. No PACS, já estão disponíveis aproximadamente 10.000 exames de ecocardiografia e mais que 6.000 exames de retina. O seguimento anual dos participantes para monitoramento de desfechos está sendo realizado no sistema em todos os CIs. A experiência adquirida permitiu o aperfeiçoamento do sistema *web* para sua utilização mais efetiva em futuras ondas.

### **AGRADECIMENTO**

Ao Dr. Lloyd Chambliss (University of North Carolina, Chapel Hill) pelo apoio e orientação dados nos últimos anos ao longo do processo de criação do Centro de Dados do ELSA.

<sup>h</sup> ELSA-Brasil. [citado 2013 mar 4]. Disponível em: [www.elsa.org.br](http://www.elsa.org.br)

## REFERÊNCIAS

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 (IL).
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.
8. Rev Saude Publica. 2001;35(4):321-40. DOI:10.1590/S0034-89102001000400001
9. Kent B, Andres C. Extreme Programming explained: embrace change. Boston (MA): Addison-Wesley; 2000.
10. Malleri 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.
11. 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:10.1191/1740774505cn63oa
12. 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
13. Schwaber K. Agile project management with scrum. Redmond (WA): Microsoft Press; 2004.
14. Winget M, Kincaid H, Lin P, Li L, Kelly S, Thornquist M. A web-based system for managing and co-ordinating multiple multisite studies. *Clinical Trials.* 2005;2(1):42-9. DOI:10.1191/1740774505cn62oa

O Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil) foi financiado pelo Ministério da Saúde (Decit – Departamento de Ciência e Tecnologia) e Ministério de Ciência e Tecnologia (Finep – Financiadora de Estudos e Projetos e CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – Processos N°s 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).

Os autores declaram não haver conflito de interesses.

Artigo submetido ao processo de julgamento por pares adotado para qualquer outro manuscrito submetido a este periódico, com anonimato garantido entre autores e revisores. Editores e revisores declaram não haver conflito de interesses que pudesse afetar o processo de julgamento do artigo.