

Implantação de um Sistema Arquivístico Unificado.

Anderson Vidart Roman, Daniel Luis Notari

Área do Conhecimento de Ciências Exatas e Engenharias
Campus Universitário da Região dos Vinhedos (CARVI)
Universidade de Caxias do Sul (UCS)
Alameda João Dal Sasso, 800 – 95705-266 – Bento Gonçalves – RS – Brasil

{anderson_roman@hotmail.com, dlnotari@ucs.br}

Resumo. *O processo de migração de dados é importante para colaborar com a atualização de aplicações devido as constantes evoluções no setor de tecnologia da informação. A migração de dados também pode ser feita para garantir maior segurança dos dados, evitando a perda de informações em consequência de erro nos aplicativos ou erro humano. Os novos bancos de dados fazem backup constante, sem ocupar muito espaço. A preocupação com a integridade corresponde à preservação da precisão, consistência e confiabilidade das informações e sistemas pela empresa ao longo dos processos ou de seu ciclo de vida. É importante que os dados circulem ou sejam armazenados do mesmo modo como foram criados, sem que haja interferência externa para corrompê-los, comprometê-los ou danificá-los. Logo, algumas técnicas são fundamentais para garantir a integridade e a coerência das informações, sem a necessidade da revisão e alteração dos registros anteriormente prescritos. O presente trabalho teve por objetivo migrar os dados de sistemas legados. Os resultados serão visualizados, analisados e validados pelo Instituto Memória Histórica e Cultural (IMHC), localizado na Universidade de Caxias do Sul (UCS).*

1. Introdução

A sociedade passa por um período de geração de grandes volumes de dados. Este volume gera uma certa preocupação na forma de tratar e organizar estes dados, sejam elas textuais ou gráficas. Segundo o Conselho Nacional de Arquivos CONARQ (Brasil, 2016) “a degradação e a obsolescência tecnológica são os principais fatores de comprometimento da preservação dos documentos digitais, uma vez que ameaçam sua autenticidade, integridade e acessibilidade”.

Arquivologia, segundo o CONARQ (Brasil, 2016) é a “disciplina que estuda as funções do arquivo e os princípios e técnicas a serem observados na produção, organização, guarda, preservação e utilização dos arquivos”.

Com a grande evolução dos sistemas, segundo Pinto e Braga (2005), há também uma necessidade maior de compreensão de um programa quando surge a necessidade de mudanças. Especialmente quando nos referimos a sistemas legados, eles são difíceis de serem modificados e compreendidos. Para facilitar a manutenção de um sistema legado, que ao longo do tempo pode ter tido sua estrutura inicial comprometida, é necessário aplicar a reengenharia, sempre visando a melhoria da estrutura e sua inteligibilidade. A reengenharia envolve a redocumentação do sistema, alterações de linguagem de programação para uma mais atual, refatoração da arquitetura e uma série de mudanças e atualizações de sua estrutura e seus dados. (PAIVA, L; SANTANDER, V., 2004).

Para Pressman (2016), a reengenharia se preocupa em analisar o projeto, em implementar um sistema legado, na aplicação das técnicas, métodos e transformação, visando uma melhora na arquitetura do sistema. Segundo Souza e Arakaki (2004), qualquer mudança deve ser feita com muita prudência, pois a transição deve ocorrer de forma transparente e sem perda de dados, visando um resultado satisfatório.

Devido ao grande volume e importância das informações, segundo Harris (2010), a migração de uma base de dados antiga para uma nova, deve ser vista como prioridade para a conclusão de um projeto que visa a substituição de *software*. A migração de dados é o processo de importação de dados legados para o seu novo sistema. Ela pode envolver a reinserção manual dos dados ou a importação deles por meio de aplicativos de migração de terceiros ou outros métodos. Para realizar um projeto deste tipo, as ações a serem realizadas dependem da escolha inteligente de uma estratégia de migração, basicamente, as duas estratégias mais discutidas na literatura são o *Big Bang* e o *Trickle* (HELENO R. D. M, 2009; Ramalho et al., 2012).

O Instituto Memória Histórica e Cultural (IMHC), possui um conjunto de ferramentas e sistemas implantados para catalogação, armazenamento e preservação dos dados, com o passar do tempo identificou-se a necessidade de implantar um sistema arquivístico unificado no IMHC.

O sistema legado Gallery, é responsável pelo gerenciamento de imagens da instituição IMHC. A leitura e análise destes dados pode ser realizada através do próprio SGDB MySQL, através da ferramenta PHP, os arquivos foram disponibilizados a partir de um *backup* da base de dados original disponibilizado previamente.

Devido a heterogeneidade dos bancos de dados e diferentes sistemas de gerenciamento de documentos dentro do instituto, que está localizado na Universidade de Caxias do Sul (UCS), e, vinculado ao Centro de Ciências Humanas e da Educação (CCHE), é necessário unificar os dados da ferramenta chamada Gallery, que até então, é responsável pela catalogação e gerenciamento de imagens. A ferramenta AtoM, possui um banco de dados totalmente diferente, mas tem a mesma ideia da ferramenta legada. Posteriormente, se encerrará qualquer cadastro feito nesta ferramenta legada e o ICA AtoM será adotado como sistema único e principal.

O sistema ICA AtoM (acrônimo para *Acess to Memory*) é totalmente feito e preparado para a ambiente *web*, é um *software* de código livre projetado para descrições arquivísticas, possui acesso a vários idiomas e tem compatibilidade com todos os navegadores existentes atualmente. Destinado a auxiliar nos processos de descrição e catalogação arquivísticas conforme os padrões estipulados pelo ICA.

O objetivo geral deste trabalho é migrar dados do sistema legado Gallery e realizar a importação no AtoM. O presente artigo apresenta na seção 2 os assuntos sobre sistema arquivístico, sistema legado, ETL, migração de dados, trabalhos relacionados e as considerações finais. Já na seção 3, descreve o que é o IMHC, os *softwares* utilizados, problemas existentes, o ambiente de teste de *software*, dados utilizados nos testes e a metodologia de integração dos dados. A seção 4 apresenta os resultados do estudo, e por fim, na seção 5, é apresentada a conclusão do trabalho.

2. Referencial Teórico

2.1. Sistema Arquivístico

A grande quantidade de informação digital trouxe uma nova realidade para arquivos. Grandes empresas e entidades passaram a adotar sistemas para gerir seus documentos. Conforme Lopes (2009, p. 359) “vive-se em uma nova etapa, em que a presença da informação registrada mensurável em *bytes*, tem uma significação muito maior do se possa, a priori, imaginar”. Neste contexto, estes sistemas de gestão e arquivamento de documentos fornecem dados de acervo, organizando e tornando possível seu gerenciamento.

Toda informação que transita pela rede, com origem de sistemas digitais, possuem uma característica arquivística. Conforme afirma o autor Lopes (2009, p. 360) “os equipamentos de informática são usados rotineiramente como meios de produzir, acumular e transmitir informações arquivística e não-arquivísticas”. Ainda segundo o autor, grande parte desta documentação é informação de arquivo, muitas destas informações “se cristalizam como documentos informáticos, não existindo em outros suportes” (LOPES, 2009, p. 360).

Para Alves (2010), com a necessidade de estabelecer normas ou regras, culminou-se na construção de catálogos, originando assim a catalogação. Com o tempo foi aprimorada e se tornou uma metodologia para processamento e tratamento de conhecimento. Com o passar do tempo e em sua respectiva época, a tecnologia auxiliou e aprimorou a representação das informações, facilitando a recuperação e disseminação da informação.

A função dos arquivos estaria na perspectiva das possíveis reutilizações da informação gerada e estruturada por processos de trabalho, que lhes impõem uma interpretação contextual. Logo, é possível expressar o conjunto da informação orgânica textual em um ambiente seguro. Pode-se considerar toda e qualquer informação gerada por instituições que disponibilizam arquivos e dados como uma informação arquivística, trazendo consigo, informações das respectivas estruturas de organização do acervo. Afinal, o uso da internet facilita a divulgação de acervos o “uso deste meio para a difusão de informações, em especial a descrição arquivística” (LOPES, 2009, p. 362).

Na próxima seção é apresentado um conceito do que é um sistema legado, seus problemas e como são vistos pelas organizações.

2.2. Sistema Legado

Pressman (2016), define como sistema legado um sistema antigo, mal documentado e projetado, que precisa ser mantido por muitos anos pelas organizações. Com o passar dos anos, este sistema se torna obsoleto, tornando vital a substituição do mesmo para que as organizações sigam com suas atividades normalmente.

Com o passar do tempo, sistemas se tornam maiores e com um nível de complexibilidade mais elevado, atendendo a muito mais colaboradores, com aumento de requisitos e dados que precisam ser alterados frequentemente. Porém com todas essas mudanças, o cenário fica desfavorável quando é preciso criar um produto viável e colocá-lo o mais rápido possível em funcionamento (FOWLER, 2014).

É quase que unânime entre as organizações optar por prolongar a vida útil dos sistemas sendo que o mesmo ainda atenda às necessidades do negócio. Muitos desses sistemas legados ainda são pilares para as empresas. Dependem em tempo integral dos serviços que são fornecidos pelos softwares, onde qualquer falha provoca um efeito direto e negativo na rotina de trabalho (SOMMERVILLE, 2016). Pinto (2005), define que a evolução de um sistema é um termo amplo que cobre todo o tempo entre uma simples adição de um campo a um banco de dados até a completa reimplantação de um sistema.

Na próxima seção são apresentados os conceitos e etapas de ETL.

2.3. ETL

Extração, transformação e carga do sistema (ETL), consomem grande parte de tempo e precisam de muito esforço para que o ambiente de *Data Warehouse* (DW) seja construído. O desenvolvimento de um sistema ETL é um grande desafio, pois diversos fatores externos acabam gerando uma enorme pressão, podemos cita-los seno os requisitos de negócio, fonte de dados, janelas orçamento, processamento e o conjunto de habilidades das pessoas envolvidas. No entanto, pode ser difícil entender o porquê do sistema ETL ser tão complexo e consumir muitos recursos (KIMBALL; ROSS, 2013).

Muitas empresas possuem uma grande variedade de ambientes operacionais onde ficam localizados seus sistemas. Conseguir unir toda essa quantidade de informações de ambientes e padrões diferentes sem que os dados sejam redundantes se torna uma tarefa árdua (KIMBALL; ROSS, 2013). Conforme descrito por Kimbal e Ross (2013), precisa-cumprir várias fases para desenvolvimento de um ETL que atende o DW.

Segundo Kimball e Caserta (2004), as empresas têm apostado em sistemas de apoio a decisão (SAD), baseada em arquiteturas DW (*Data Warehouse*) para melhor gerir essa grande quantidade de dados. Porém para que tudo isso seja possível é obrigatório ter um processo de ETL. Os autores ainda afirmam, 70% dos recursos necessários para a implementação e manutenção de um DW são consumidos neste processo.

Segundo Kimball e Caserta (2004, p. 22), um sistema de ETL refere-se à extração, remoção dos erros e correção dos dados em falta, fornece medidas documentadas de confiança nos dados, captura o fluxo de dados transacionais por precaução, ajusta dados de várias fontes para ser utilizado em conjunto e estrutura os dados para serem utilizados por ferramentas de utilizadores finais. Um sistema ETL segundo Kozielski e Wrembel (2008, p. 21) é composto por três (3) etapas ou fases, a primeira delas é a extração, onde é realizada a extração dos dados heterogêneos em diversas fontes, a segunda consiste na transformação dos dados, onde é realizada a limpeza, padronização dos arquivos, etc., e pôr fim na terceira os dados transformados são inseridos em vários Banco de Dados (BD) de destino de DW.

Seguindo os requisitos, identificamos uma série de decisões de arquitetura que você precisa fazer no início do seu projeto de ETL (KIMBALL; CASERTA, 2004, p. 38). Segundo Kimball (2013), alguns itens precisam estar bem alinhados antes de iniciar um projeto de ETL, os requisitos de negócio precisam estar claros e documentados, é necessário realizar uma análise de viabilidade dos dados, é preciso verificar a latência dos dados, ou seja, qual é o tempo máximo para que os dados sejam disponibilizados e quais são as políticas de conformidade e segurança adotadas pela empresa.

A próxima seção tem por objetivo apresentar conceitos sobre a migração de dados, relacionados ao objetivo deste artigo.

2.4. Migração de Dados

As diferentes tecnologias para gerenciamento de dados e constantes avanços tecnológicos, faz com que em um determinado momento o que era moderno e eficiente se torne obsoleto. Nesse momento a migração de dados para um novo sistema se faz necessário e é tomada pela maioria dos gerentes de projeto, tendo como finalidade atender as necessidades atuais e futuras da empresa (STAIR; REYNOLDS, 2005).

A migração de um sistema legado para um novo, independentemente do tipo de projeto, necessita alguns cuidados, pois conversões, tratamentos ou transformações dos dados serão necessárias para que o processo de migração seja realizado com sucesso sem perda de informações (HARRIS, 2010).

A migração de dados permite que dados sejam transformados e se tornem compatíveis para serem inseridos em um novo sistema, porém esse processo de migração deve ser realizado com muita competência e cuidado. É muito importante que a migração seja feita com a melhor qualidade e integridade, preservando os dados e garantindo a continuidade de operação da empresa. Para garantir que nenhuma informação seja perdida no processo e seja realizada de forma inteligente, existem duas estratégias conhecidas, *Big Bang e Trickle* (MISTRY, 2014).

A primeira estratégia *Big Bang*, determina que a migração deve ser realizada somente uma vez, o sistema que está passando pelo processo de migração deve ficar inacessível, enquanto todo o processo de ETL é realizado. Para minimizar os impactos negativos deste tipo de estratégia, a maioria das empresas realizam este processo após o horário de trabalho ou em feriados longos, desta maneira o corte ao acesso dos dados não impacta o fluxo vital da organização. A segunda estratégia *Trickle*, determina que a migração deve ser realizada de forma iterativa, ou seja, diz-se do processo que se repete diversas vezes para se chegar a um resultado e a cada vez gera um resultado parcial que será usado na vez seguinte, até que não tenham mais dados a ser migrados. Desta maneira o processo se torna muito mais complexo pois existirão dois sistemas rodando em paralelo, onde os usuários iram precisar alternar entre um sistema e outro para localizar as informações desejadas, até que o processo de migração não seja finalizado os usuários precisam continuar com acesso e utilizar o sistema legado (HARRIS, 2010).

Na próxima seção são apresentados os trabalhos relacionados a implantação de um sistema arquivístico unificado.

2.5. Trabalhos Relacionados

Para o desenvolvimento desta etapa, revisou-se o conteúdo disponível em uma base de dados eletrônica, no período entre 2003 e 2017. Pode-se destacar alguns deles como uma base de dados confiável, gratuita e que disponibilizasse o material integralmente.

Após iniciou-se a seleção de trabalhos científicos cujo títulos e/ou resumos fossem relacionadas com as palavras chaves, Sistema Arquivístico, Sistema Legado, *Data Warehouse*, ETL e Migração de Dados. O quadro 1 descreve os artigos que foram selecionados para análise, apresentando o título e os autores.

Quadro 1 – Artigos selecionados

ID	Título	Autores
A1	Acervos Audiovisuais em Brasília: Imagem, Memória e Informação	Miriam Paula Manini (2015)
A2	Ambiente de <i>Data Warehousing</i> para Integração de Dados de Saúde Pública em Âmbito de Gestão Regional	Samuel Zanferdini Oliva (2015)
A3	Aplicação de Técnicas de ETL para a Integração de dados com Ênfase em <i>Big Data</i> na Área de Saúde Pública	Clícia dos Santos Pinto (2015)
A4	Conhecendo Sistemas Legados através de Métricas de <i>Software</i>	Cristiane Soares Ramos (2004)
A5	<i>Data Warehouse</i> : O processo de Migração de Dados	Laura Costa Sarkis (2001)
A6	Fases para o Desenvolvimento de uma Extração Transformação e Carga	Carles Mateus Manzoli Silva, Richard Ribeiro de Carvalho e Rodrigo Vitorino Moravia (2015)
A7	Gestão da Preservação de Documentos Arquivísticos Digitais: Proposta de um Modelo Conceitual	Humberto Celeste Innarelli (2015)
A8	Gestão de Documentos Arquivísticos Digitais: Análise do Sistema ELB	Roberta Elias da Rocha (2016)
A9	Migração de Dados: Sistemas de Informações Aplicados na Integração Automática de Bancos de Dados / Migração entre bases de dados heterogêneas: um estudo de caso em um instituto de história cultural	Giovani Delinger (2017)
A10	O Processo ETL: O Caso da Unitel + Telecomunicações	Edmir de Jesus Oliveira Tavares (2013)
A11	Ontologia Digital Arquivística: Interoperabilidade e Preservação da Informação Arquivística em Sistemas Informatizados de Arquivos e na <i>Web</i>	Charlley dos Santos Luz (2016)
A12	Requisitos para Ferramentas de Migração de Dados	Pedro de Alcântara dos Santos Neto, Josino Rodrigues Neto, Francisco das Chagas Ribeiro Júnior e Pedro Almir Oliveira (2012)
A13	Tecnologia da informação aplicada no gerenciamento da informação histórico cultural da Serra Gaúcha	Ariel Felippi (2016)
A14	Um Processo de Migração de Sistema Legado Funcional para Orientado a Objetos Direcionado por Indicadores de Qualidade	Wagner Leal dos Santos (2007)

Fonte: Elaborado pelo autor.

O quadro 2 apresenta o resumo dos artigos que foram selecionados para análise, apresentando o título e objetivo dos autores.

Quadro 2 – Quadro resumo

(continua)

ID	Título	Objetivo
A1	Acervos Audiovisuais em Brasília: Imagem, Memória e Informação	Construção de um mapa, a fim de, identificar e quantificar os itens da produção e acumulação de registros audiovisuais.
A2	Ambiente de <i>Data Warehousing</i> para Integração de Dados de Saúde Pública em Âmbito de Gestão Regional	Utilização de um ambiente de Data Warehouse a fim de viabilizar de forma otimizada e simples o monitoramento e a análise de informações na área da saúde pública.
A3	Aplicação de Técnicas de ETL para a Integração de dados com Ênfase em <i>Big Data</i> na Área de Saúde Pública	Utilização de técnicas de ETL no desenvolvimento de pré-processamento para o pareamento probabilístico de registros em bases de dados na área da saúde pública.
A4	Conhecendo Sistemas Legados através de Métricas de <i>Software</i>	Apresentação de um modelo de métricas, definido através da utilização e abordagem Goal-Question-Metric para apoiar o melhor entendimento sobre um sistema legado.
A5	<i>Data Warehouse</i> : O processo de Migração de Dados	Estudo do processo de migração de dados para o Data Warehouse.
A6	Fases para o Desenvolvimento de uma Extração Transformação e Carga	Auxiliar na escolha de um sistema de apoio à tomada de decisão de uma organização, além de abordar o cuidado com os procedimentos técnicos e gerenciais que a solução deve ter.
A7	Gestão da Preservação de Documentos Arquivísticos Digitais: Proposta de um Modelo Conceitual	Identificar as variáveis que impactam diretamente as políticas de preservação dos documentos arquivísticos digitais.
A8	Gestão de Documentos Arquivísticos Digitais: Análise do Sistema ELB	Evidenciar a importância da gestão de documentos digitais para identificar e gerenciar os documentos arquivísticos digitais a fim de manter suas características, especialmente a autenticidade e a organicidade.
A9	Migração de Dados: Sistemas de Informações Aplicados na Integração Automática de Bancos de Dados / Migração entre bases de dados heterogêneas: um estudo de caso em um instituto de história cultural	Aplicar os sistemas de informações no tratamento e migração dos dados dos sistemas legados do IMHC. Visando exportar as informações para uma nova estrutura de dados definida pela aplicação implementada.
A10	O Processo ETL: O Caso da Unitel + Telecomunicações	Compreender os principais mecanismos técnicos, ferramentas do processo ETL e as ferramentas associadas.
A11	Ontologia Digital Arquivística: Interoperabilidade e Preservação da Informação Arquivística em Sistemas Informatizados de Arquivos e na <i>Web</i>	Verificar a aplicação de ontologias criadas para possibilitar a manutenção da cadeia de custódia e do contexto da informação orgânica e do documento arquivísticos.
A12	Requisitos para Ferramentas de Migração de Dados	Apresentar um conjunto de requisitos relacionados as ferramentas para migração de dados e um estudo e comparativo de ferramentas de migração de dados.
A13	Tecnologia da informação aplicada no gerenciamento da informação histórico cultural da Serra Gaúcha	Implantação de softwares de código aberto em instituições de preservação da informação histórica cultural.

(conclusão)

A14	Um Processo de Migração de Sistema Legado Funcional para Orientado a Objetos Direcionado por Indicadores de Qualidade	Propor a organização de práticas e técnicas de engenharia de software que permita a evolução tecnológica do software com redução do tempo no desenvolvimento e melhoria da qualidade.
------------	---	---

Fonte: Elaborado pelo autor.

No projeto de Miriam Paula Manini (2015) foi apresentado um estudo sobre um projeto que está mapeando a produção e a acumulação de registros audiovisuais. O objetivo deste projeto foi a construção de um mapa, a fim de, identificar e quantificar os itens da produção e acumulação de registros audiovisuais.

A dissertação de Pós-graduação de Samuel Zanferdini Oliva (2015) visou organizar e disponibilizar o conhecimento sobre informações da saúde combinado com aspectos estratégicos e com a tecnologia da informação. O autor propôs a utilização de um ambiente de *Data Warehouse* a fim de viabilizar de forma otimizada e simples o monitoramento e a análise de informações na área da saúde pública. Visando que uma tomada de decisão seja mais ágil e precisa por parte dos gestores, aperfeiçoando o atendimento à saúde para a população.

A dissertação de Mestrado de Clícia dos Santos Pinto (2015) teve como proposta a utilização de técnicas de ETL no desenvolvimento de um módulo de pré-processamento para o pareamento probabilístico de registros em bases de dados na área da saúde pública. A autora utilizou a ferramenta de processamento distribuído do Spark, para garantir o tratamento adequado para o contexto de *Big Data* na qual sua pesquisa está inserida, gerando assim, respostas em tempo hábil.

O artigo de Cristiane Soares Ramos (2004) apresentou um modelo de métricas, definido através da utilização e abordagem *Goal-Question-Metric* para apoiar o melhor entendimento sobre um sistema legado. No que se refere a seu código e documentação fornecendo informações para apoiar uma empresa na definição de seus contratos de manutenção de *software*.

A dissertação de Pós-graduação de Laura Costa Sarkis (2001) descreve os conceitos básicos do ambiente do *Data Warehouse*, abordando em especial o processo de migração de dados. O objetivo geral foi o estudo do processo de migração de dados para o *Data Warehouse*, através da descrição e análise do conteúdo de quatro abordagens. Sendo elas, uma arquitetura de extração de dados, um plano de conversão de dados e especificações para abordar o processo de migração e garantir a qualidade de dados, estratégias de migração de dados através dos processos de perfilamento e mapeamento dos dados e a utilização de tecnologias de movimento de dados na preparação de dados para o *Data Warehouse*.

O artigo de Carles Mateus Manzoli Silva, Richard Ribeiro de Carvalho e Rodrigo Vitorino Moravia (2015) é um estudo para o levantamento de fases no desenvolvimento de ETL. Etapa constituinte na construção de um *Data Warehouse* (DW), explicando as fases da etapa de ETL apontando as melhores práticas e o fluxo preferencial dos projetos de implantação dos sistemas de apoio à tomada de decisão. O objetivo foi auxiliar na escolha de um sistema de apoio à tomada de decisão de uma organização, além de abordar o cuidado com os procedimentos técnicos e gerenciais que a solução deve ter, contando sempre com atenção por parte dos profissionais de Tecnologia da Informação.

A tese de Pós-graduação de Humberto Celeste Innarelli (2015) buscou apresentar os problemas da preservação de documentos arquivísticos digitais frente as teorias e práticas arquivísticas e as tecnologias de informação e da comunicação. O objetivo desta tese é identificar as variáveis que impactam diretamente as políticas de preservação dos documentos arquivísticos digitais.

O trabalho de conclusão de curso de Roberta Elias da Rocha (2016) teve como objetivo evidenciar a importância da gestão de documentos digitais para identificar e gerenciar os documentos arquivísticos digitais a fim de manter suas características, especialmente a autenticidade e a organicidade. A pesquisa focaliza a fase de produção do documento, especialmente o momento de sua elaboração. Ressalta ainda a importância da gestão de documentos digitais para identificar e gerenciar os documentos arquivísticos a fim de manter suas características especialmente a autenticidade e a organicidade.

O trabalho de conclusão de curso de Giovani Delinger (2017) apresentou um estudo sobre integração de bancos de dados, realizando um estudo de caso real, no Instituto Memória Histórica e Cultural (IMHC) da Universidade de Caxias do Sul (UCS). O autor visou aplicar os sistemas de informações no tratamento e migração de dados de um sistema legado do IMHC, pesquisando e estudando ferramentas para popular a base de dados realizando a migração destas informações para um novo sistema do IMHC. O artigo de Giovani Delinger (2017) buscou aplicar os sistemas de informações no tratamento e migração dos dados dos sistemas legados do IMHC. Visando exportar as informações para uma nova estrutura de dados definida pela aplicação implementada. Realizou ainda um estudo para identificar, técnicas e métodos que possibilitam o resguardo do retrabalho manual dos colaboradores do instituto. O processo do autor é realizado em etapas, onde ele realiza a configuração do laboratório de testes, tornando possível analisar o comportamento do sistema, buscando compreender o funcionamento de tratamento e armazenagem dos dados. Realiza o estudo de ferramentas para auxiliar no processo de migração das informações, identifica possíveis falhas e pontos críticos no processo de importação e verifica se os registros importados na interface foram gravados com sucesso.

A monografia do autor Edmir de Jesus Oliveira Tavares (2013) visou compreender os principais mecanismos técnicos, ferramentas do processo ETL e as ferramentas associadas. Apresenta um enquadramento teórico sobre esse processo e onde são distinguidas as principais etapas de ETL. Cita também as ferramentas de ETL (Comerciais e *OpenSource*), *Talend Open Studio for Data Integration*, uma ferramenta utilizada para implementação de sistemas de ETL e realiza um estudo de caso prático sobre esses sistemas.

A dissertação de Charley dos Santos Luz (2016) teve como objetivo verificar a aplicação de ontologias criadas para possibilitar a manutenção da cadeia de custódia e do contexto da informação orgânica e do documento arquivístico. Visou facilitar sua interoperabilidade para uso na difusão de suas informações arquivísticas e na preservação digital a longo prazo, em sistemas de arquivo quanto na rede mundial de computadores.

O artigo de Pedro de Alcântara dos Santos Neto, Josino Rodrigues Neto, Francisco das Chagas Ribeiro Júnior e Pedro Almir Oliveira (2012) apresentou um conjunto de requisitos relacionados as ferramentas para migração de dados. Os autores também realizam um estudo e comparativo de ferramentas. A finalidade desse estudo é verificar

o nível de atendimento aos requisitos propostos, além de servir como base para avaliação das ferramentas pesquisadas.

O artigo de Ariel Felippi (2016) teve como objetivo a implantação de *softwares* de código aberto em instituições de preservação da informação histórica cultural. O autor realizou um estudo de caso das ferramentas existentes no acervo histórico cultural do IMHC. Os resultados indicaram que os *softwares* proporcionarão ao IMHC as funcionalidades necessárias para o gerenciamento do acervo arquivístico. Por fim, analisaram ferramentas de Tecnologia da Informação e Comunicação que pudessem auxiliar no processo de implantação.

A dissertação de Wagner Leal dos Santos (2007) tem como principal objetivo, propor a organização de práticas e técnicas de engenharia de *software*, um processo aplicável que permita a evolução tecnológica do *software* com redução do tempo no desenvolvimento e melhoria da qualidade.

Na próxima seção são apresentadas as considerações finais dos trabalhos relacionados a implantação de um sistema arquivístico unificado.

2.6. Considerações Finais

Analisando os trabalhos relacionados, destaco dois deles que foram de grande importância para concretizar este trabalho. O artigo de Ariel Felippi (2016) auxiliou no entendimento sobre a implantação de *softwares* de código aberto. Na escolha de *softwares* com as funcionalidades necessárias para o gerenciamento do acervo arquivístico, afim de auxiliara o processo de implantação.

O trabalho de conclusão de curso de Giovani Delinger (2017), auxiliou no tratamento e migração dos dados dos sistemas legados do IMHC. Também se fez necessário para entender a forma de exportar as informações para uma nova estrutura de dados e para identificar, técnicas e métodos que possibilitam o resguardo do retrabalho manual dos colaboradores do instituto.

Na próxima seção são apresentadas as metodologias utilizadas para a implantação de um sistema arquivístico unificado.

3. Metodologia

3.1. Instituto Memória Histórica e Cultural (IMHC)

O Instituto Memória Histórica e Cultural, da Universidade de Caxias do Sul, atua em cinco áreas relacionadas com a preservação e o estudo da memória, individual e coletiva, pessoal e institucional, nos diversos campos da cultura. Sua perspectiva é rigorosamente interdisciplinar, tendo como focos principais de pesquisa e de ação o patrimônio cultural, tanto material como imaterial, na sua dinâmica histórico-social.¹

A atuação do IMHC (Instituto Memória Histórica e Cultural) contempla quatro níveis de trabalho, articulados entre si, a coleta, organização, processamento e preservação do acervo histórico da Instituição, da Cultura Regional e outros considerados de importância histórica e cultural. O levantamento e registro de dados e fenômenos de interesses histórico e cultural, constituindo-se, a partir deles, acervos documentais e

¹ <https://www.ucs.br/site/instituto-memoria-historica-e-cultural/>

bancos de dados. A análise e a interpretação dos dados, em busca de possíveis significados históricos e culturais e atividades de transparência, para a sociedade, do conhecimento produzido e organizado, sob a forma de cursos, consultoria, assessoria e outros serviços contratados mediante assinatura de convênios, para instituições públicas e privadas através de atividades. São elas, prospecção para a criação de museus, arquivos e memoriais, diagnóstico de museus, arquivos e memoriais (com orçamentos), organização e gestão de acervos, pesquisa e publicações institucionais, criação de programas de divulgação e de ação educativa, treinamentos (cursos, oficinas, palestras, etc.) para professores e funcionários de museus, centros de documentação e arquivos, desenvolvimento de atividades de educação patrimonial voltadas para o ensino em suas diversas instâncias, desde o ensino fundamental até a pós-graduação e caracterização de acervo urbano para processos de planejamento e preservação.

O IMHC compõe-se da seguinte estrutura: o LEPAR (Laboratório de Estudo de Arqueologia), o Programa ECIRS (Elementos culturais das antigas colônias italianas da Região Nordeste do Rio Grande do Sul), Programa IRIS (Investigação e Resgate de Imagem e Som), Centro de Documentação da Universidade de Caxias do Sul - CEDOC/UCS, e o CMRJU (Centro de Memória Regional do Judiciário). É também órgão de apoio e de suporte ao Programa de Pós-Graduação em Letras, Cultura e Regionalidade, ao Programa de Pós-Graduação em Educação, ao Programa de Pós-Graduação em História e ao Programa de Pós-Graduação em Turismo, todos no nível de mestrado.

Na próxima seção são apresentados os *softwares* presentes no IMHC.

3.2. Softwares Utilizados

O IMHC, possui um conjunto de ferramentas e sistemas implantados para catalogação, armazenamento e preservação dos dados, com o passar do tempo identificou-se a necessidade de implantar um sistema arquivístico unificado no IMHC.

O sistema legado Gallery, é responsável pelo gerenciamento de imagens da instituição IMHC. A leitura e análise destes dados pode ser realizada através do próprio SGDB MySQL, através da ferramenta PHP, os arquivos foram disponibilizados a partir de um *backup* da base de dados original disponibilizado previamente.

Devido a heterogeneidade dos bancos de dados e diferentes sistemas de gerenciamento de documentos dentro do instituto, que está localizado na Universidade de Caxias do Sul (UCS), e, vinculado ao Centro de Ciências Humanas e da Educação (CCHE), é necessário unificar os dados da ferramenta chamada Gallery, que até então, é responsável pela catalogação e gerenciamento de imagens. A ferramenta AtoM, possui um banco de dados totalmente diferente, mas tem a mesma ideia da ferramenta legada. Posteriormente, se encerrará qualquer cadastro feito nesta ferramenta legada e o ICA AtoM será adotado como sistema único e principal.

O sistema ICA AtoM (acrônimo para *Acess to Memory*) é totalmente feito e preparado para a ambiente *web*, é um *software* de código livre projetado para descrições arquivísticas, possui acesso a vários idiomas e tem compatibilidade com todos os navegadores existentes atualmente. Destinado a auxiliar nos processos de descrição e catalogação arquivísticas conforme os padrões estipulados pelo ICA.

Na próxima seção são apresentados os problemas existentes ao decorrer do processo de migração.

3.3. Problemas Existentes

A importação da base de dados do sistema Gallery, originou-se por meio de uma exportação do banco de dados MARIADB, pertinente ao sistema de gerenciamento de imagens Gallery. Deste modo, se observou que o banco de dados ATOM, encontra-se estruturado por meio de chaves estrangeiras. Em vista disso, se fez necessário uma análise mais profunda buscando identificar o funcionamento do banco em si.

Segundo Fellipe (2016), observou que o AtoM trabalha com uma estrutura complexa para o gerenciamento dos arquivos. Para cada upload realizado, o *software* analisa qual o grau de descrição arquivística daquele arquivo; se o mesmo pertence a alguma descrição arquivística já existente; qual o autor do arquivo; qual a instituição arquivística detentora do mesmo; sob qual regra de descrição arquivística o arquivo está sendo salvo; se o mesmo está classificado como multi-repositório (compartilhado entre atores ou instituições), entre outras. Foi necessário analisar o código fonte da aplicação AtoM em busca da biblioteca responsável pela realização dos uploads, comentá-la e após efetuar a simulação de envio dos arquivos, para que o próprio AtoM realize o devido gerenciamento dos documentos arquivísticos e as inserções no banco de dados.

Segundo Delinger (2017), a dificuldade na obtenção dos dados para o layout proposto pela Artefactual foi que o sistema legado da instituição não tinha padronização alguma, contendo basicamente uma descrição da publicação e informações do objeto digital vinculado, podendo ser capturadas de uma única entidade ao contrário do AtoM, que cria objetos derivativos a partir de cada objeto principal. Apesar da desenvolvedora Artefactual, informar na documentação do AtoM que a ordem das colunas nos modelos de arquivos CSV é a mesma que a ordem na interface da aplicação, e não deveriam ser alteradas, a personalização das colunas foi necessária. Diversas colunas não puderam ser preenchidas por falta de informação do sistema legado da instituição.

Na próxima seção é apresentado o ambiente de teste de *software* definido para atender a necessidade do IMHC.

3.4. Ambiente de Teste de Software

O ambiente de testes foi definido para atender as necessidades do Instituto diante do modelo tecnológico utilizado pelo mesmo no gerenciamento e preservação da informação histórica cultural.

Ao iniciar o processo de análise do comportamento do sistema AtoM, foi realizada a instalação e configuração de uma máquina virtual com as especificações técnicas similares a do servidor de produção, que possui as seguintes características de processador Intel Xeon Octa Core 2394 MHz, disponibilizado para o presente trabalho 2 Core, 2040 MB (2GB) de memória RAM, Hard Disk (HD) de 50 GB, sistema operacional (S.O.) Red Hat Enterprise Linux Server versão 7.3 com arquitetura de 64bits.

Na etapa seguinte, foi iniciado o processo de instalação das dependências do *software*, como o Apache sendo o servidor *web* da aplicação, o MySQL como sistema de gerenciamento de banco de dados (SGBD), o *Hypertext Preprocessor* (PHP) como linguagem de script open source de uso geral para interpretação das páginas, e o Elasticsearch que é um servidor de pesquisa relativamente novo baseado no Apache Lucene desenvolvido em Java, que trouxe ao sistema muitos recursos avançados além do desempenho nas pesquisas e escalabilidade, e finalmente ocorreu a instalação do AtoM.

Na próxima seção são apresentadas a origem e modelo de dados utilizados nos testes.

3.5. Dados Utilizados nos Testes de Software

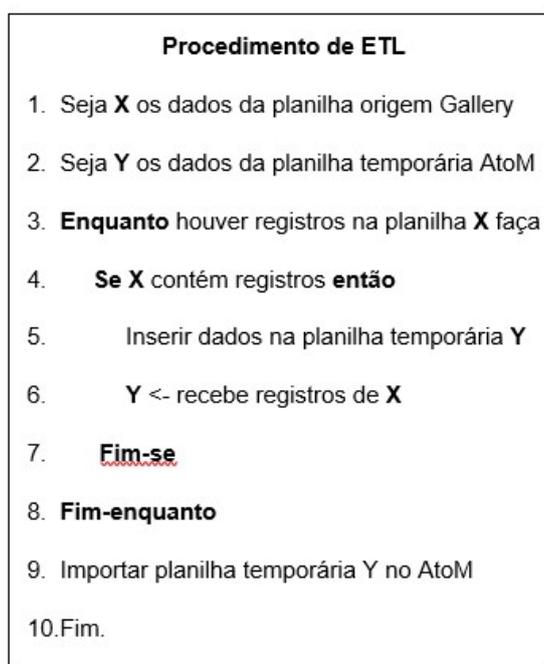
A importação da base de dados do sistema Gallery, originou-se por meio de uma exportação do banco de dados mariadb, pertinente ao sistema de gerenciamento de imagens Gallery. O arquivo SQL oriundo desta exportação, possui 117 Gigabytes (GB), o tamanho do referido arquivo tornou onerosa a usabilidade do mesmo no processo de importação junto a aplicação MySQL.

Na próxima seção são apresentados o método e as ferramentas utilizadas para a migração dos dados.

3.6. Método de Integração dos Dados

A migração de dados ocorreu através da aplicação direta de ETL, utilizando a ferramenta MySQL e Excel, no qual auxiliaram na extração, tratamento e organização dos dados do sistema legado, gerando um arquivo CSV para posteriormente realizar a importação no AtoM. O algoritmo utilizado para o procedimento de ETL deve ser executado em ordem, seguindo passos já definidos na Figura 1.

Figura 1 – Procedimento de ETL



Fonte: Elaborado pelo autor.

Estes são compostos por uma sequência de dados. O passo 1 contém todos os registros do sistema legado Gallery, armazenados em uma planilha X em formato CSV. No passo 2 ficarão armazenados em uma planilha temporária Y os registros da planilha X em formato CSV devidamente tratados. No passo 3 até que haja registros na planilha X, os dados continuarão a ser recebidos pela planilha temporária Y. Na etapa 9 todos os

registros presentes na planilha temporária Y, cujo formato também é CSV, por fim, será realizada a importação dos dados diretamente na ferramenta AtoM.

Na próxima seção são apresentados os resultados, explicando a utilização do algoritmo utilizado no procedimento ETL e a execução detalhada de cada passo.

4. Resultado

A migração de dados ocorreu através da aplicação direta de ETL, utilizando a ferramenta MySQL e Excel, no qual auxiliaram na extração, tratamento e organização dos dados do sistema legado, gerando um arquivo CSV para posteriormente realizar a importação no AtoM.

O algoritmo utilizado para o procedimento de ETL foi executado na seguinte ordem, no passo 1 todos os registros do sistema legado estão presentes em uma planilha X em formato CSV, no passo 2 criou-se uma planilha temporária Y já parametrizada para receber os registros da planilha X em formato CSV, no passo 3 até que haja registros na planilha X, eles serão recebidos pela planilha temporária Y e na etapa 9 todos os registros presente na planilha temporária Y, cujo formato também é CSV, serão importados através da própria ferramenta AtoM.

O processo de leitura dos dados oriundos do Gallery, através de consultas, tratamentos e transformações no respectivo SGDB, resultando as informações conforme exigências do template disponibilizado pela Artefactual. Concluindo o processo através da exportação da consulta diretamente no formato CSV, devendo ser tratados e importados na planilha temporária Y para posteriormente importar diretamente no AtoM. Em razão disto, é estabelecido um processo próprio de ETL, onde são realizadas as ações de migração através de scripts em linguagem SQL próprios.

Após a identificação dos dados hierárquicos do sistema Gallery, foi possível a estruturação adequada das informações no arquivo de modelo. Fazendo-se o uso das colunas "legacyId" e "parentId" a associação pai/filho se realizou. Os pais precisaram ser adicionados aos dados CSV antes de seus filhos, evitando assim, falha na interpretação do template, destacando que o sistema realiza a leitura de linha a linha. Conseqüentemente, o autor classificou os registros por nível de descrição. O nível de descrição para cada linha do modelo especificou-se em uma coluna chamada levelOfDescription. A palavra Fonds deu-se em registros pais e a Item em registros filhos subsequentes, conforme apresentados na Figura 2.

Figura 2 – Apresentação da coluna relative_path_cache do sistema legado

A	B	C	D	E	F
legacyId	parentId	identifier	title	levelOfDescription	extentAndMedium
27607		27607	30. Entral 2009 - Enc	Fonds	Álbum
27667	27607	27667	ENTRAI (2009) 0063	Item	Um item
27666	27607	27666	ENTRAI (2009) 0062	Item	Um item
27665	27607	27665	ENTRAI (2009) 0061	Item	Um item
27640	27607	27640	ENTRAI (2009) 0036	Item	Um item
27664	27607	27664	ENTRAI (2009) 0060	Item	Um item
27638	27607	27638	ENTRAI (2009) 0033	Item	Um item
27663	27607	27663	ENTRAI (2009) 0059	Item	Um item
27662	27607	27662	ENTRAI (2009) 0058	Item	Um item
27661	27607	27661	ENTRAI (2009) 0057	Item	Um item
27660	27607	27660	ENTRAI (2009) 0056	Item	Um item

Fonte: Elaborado pelo autor.

Após o mapeamento dos dados hierárquicos do sistema legado para o modelo CSV da planilha X, outros campos precisaram de uma avaliação particular. Colunas como a eventActors, eventStartDates, e eventEndDates estão relacionadas à criação de atores e eventos. No modelo de dados do AtoM, cada descrição arquivística é uma descrição de um registro, entendido como uma evidência documental oriunda de uma ação ou evento. Os eventos ligam os atores com as publicações. Tendo isto em vista, faz-se o uso dessas colunas acrescentando as datas do evento de criação ou acumulações associadas.

No entanto, estas informações no sistema legado estão descritas no mesmo campo de texto que descreve os objetos digitais, isto é, no mesmo campo foram descritas informações dos documentos, atores, datas e descrições dos eventos de forma despadronizada e informal. Em consequência de diversos colaboradores do instituto efetuarem cadastros de descrições, há casos onde estes dados são inexistentes. A coluna onde estes dados estão armazenados chama-se archivalHistory, conforme apresentados na Figura 3.

Figura 3 – Coluna archivalHistory

G
archivalHistory
Desfile e feira do XIV ENTRAÍ - Encontro das Tradições Italianas de 2009.Autoria das imagens e tratamento: Aldo Toniazzo.Classificação, c
Autoria das imagens e tratamento: Aldo Toniazzo.Classificação, descrição e indexação: Anthony Beux Tessari.
Coleção Santa Tereza - Santa Tereza, RS.Autoria das imagens e tratamento: Aldo Toniazzo.Classificação, descrição e indexação: Anthony I
Mostra Fotográfica "Bortolo Ceconello: de pedra, para sempre" (32 imagens), organizada pelo IMHC em comemoração aos 80 anos da C
Elementos Históricos e Culturais da Área da Barragem Machadinho.Autoria das imagens e tratamento: Aldo Toniazzo. Digitalização: Anth
Clique sobre a imagem para abrir o álbum
Ecirs - Elementos Culturais da Imigração Italiana no Nordeste do Rio Grande do SulInstituto Memória Histórica e Cultural - IMHC / Univer
Pesquisa e Registro do Patrimônio Histórico e Cultural Rural do Município de Caxias do Sul.Autoria das imagens e tratamento: Aldo Toni
Mostra Fotográfica - "Vozes de Ana Rech" - 2011 (12 imagens)

Fonte: Elaborado pelo autor.

Após a importação dos dados, podemos verificar um exemplo na Figura 4, como estas informações são levadas com sucesso para o AtoM.

Figura 4 – Coluna archivalHistory

Zona do contexto »	
Nome do produtor	Anthony Beux Tessari
História do arquivo	Ação educativa com alunas do magistério do Instituto Estadual de Educação Cristóvão de Mendoza. Na oportunidade, as alunas conheceram o trabalho desenvolvido pelo Instituto Memória Histórica e Cultural (IMHC) da UCS, como as atividades de preservação e confecção de embalagens de acondicionamento dos documentos históricos (conforme esta imagem). A ação educativa foi ministrada pelo funcionário do IMHC Anthony Beux Tessari. Caxias do Sul, RS, 22/nov./2014 Autoria: Leandra Verônica Pegoraro Miotto «

Fonte: Elaborado pelo autor.

Como estas informações são obrigatórias, definiu-se valores padrões para alguns campos. Considerando o exposto, o autor padrão para as descrições foi o diretor do instituto IMHC e as datas foram estabelecidas para a data da importação. Na sequência, fez-se necessário o preenchimento da coluna culture, indicando para a aplicação o idioma padrão das descrições que serão carregadas. O valor possível desta coluna é qualquer código no idioma ISO 639-1 de duas letras, dado como exemplo: “en” para inglês, “it” para italiano, etc. Com base no contexto apresentado, o valor preenchido para as descrições foi “pt”, indicando que as publicações do instituto permanecem em português. Com isso a planilha X ficou com todos os registros tratados para serem importados conforme o template disponibilizado pela Artefactual, que será a planilha temporária Y, como mostra a Figura 5, onde todos os registros da planilha X foram recebidos na planilha temporária Y, para realizar a importação no AtoM.

Figura 5 – Planilha Temporária Y

A	B	C	D	E	F	
legacyId	parentId	identifier	title	levelOfDescription	extentAndMedium	archivalHistory
27607		27607	30. Entrai 2009 - Encontro das Tradiç	Fonds	Álbum	Desfile e feira do XIV EN
20755		20755	35. Festa da Uva 2006 - Caxias do Su	Fonds	Álbum	Autoria das imagens e t
28347		28347	17. Coleção Santa Tereza, RS	Fonds	Álbum	Coleção Santa Tereza - S
28313		28313	46. Mostra Fotográfica - "Bortolo Ce	Fonds	Álbum	Mostra Fotográfica "Bo
2356		2356	05. Projeto Ecam - UHE Machadinho	Fonds	Álbum	Elementos Históricos e C
2353		2353	IMHC - Instituto Memória Histórica e	Fonds	Álbum	Clique sobre a imagem p
2354		2354	Programa ECIRS/IMHC/UCS - Acervo	Fonds	Álbum	Ecirs - Elementos Cultur
2355		2355	09. Projeto Ecamp - UHE Campos No	Fonds	Álbum	Elementos Históricos e C
5150		5150	03. Projeto Victur / URB-AL - Caxias	Fonds	Álbum	Pesquisa e Registro do F
13687		13687	47. Mostra Fotográfica - "Vozes de A	Fonds	Álbum	Mostra Fotográfica - "V

Fonte: Elaborado pelo autor.

O processo de importação da planilha temporária Y do arquivo CSV no AtoM, levou aproximadamente 1 dia e meio até ser concluído, todos os dados parametrizados anteriormente foram gravados com sucesso em seus respectivos campos, sem perda de informação.

Vale ressaltar que em consequência de diversos colaboradores do instituto efetuarem cadastros de descrições, há casos onde estes dados são inexistentes. Como estas informações são obrigatórias, definiu-se valores padrões para os campos. Considerando o autor padrão para as descrições o diretor do instituto IMHC assim como uma data aleatória estabelecida para a importação. Grande parte dos campos no modelo CSV tem sua nomenclatura de forma óbvia, facilitando a interpretação da informação necessária para cada caso. Mesmo utilizando meios automatizados para realizar a importação, as situações das publicações não puderam ser definidas como publicadas (*Published*) conforme exemplificado na documentação padrão do AtoM, sendo assim, as descrições permaneceram na situação rascunho (*Draft*), como mostra a Figura 6.

Figura 6 – Publicações em Situação *Draft*

Identity area »	
Reference code	64887
Title	13. Alunas do Magistério do Instituto de Educação Cristóvão de Mendoza - 22 nov. 2014
Level of description	Fonds
Extent and medium	Álbum
Context area »	
Name of creator	Anthony Beux Tessari
Conditions of access and use area »	
Language of material	• Portuguese
Access points »	
Name access points	• Anthony Beux Tessari (Creator)
Description control area »	
Institution Identifier	IMHC
Status	Draft
Level of detail	Partial

Fonte: Elaborado pelo autor.

O instituto IMHC cataloga, armazena, realiza a manutenção e preservação dos documentos do acervo histórico da Instituição, da Cultura Regional e outros, com importância histórica e cultural. Documentos arquivísticos em formato digital, a preservação dependerá principalmente da solução tecnológica adotada e dos custos que ela envolve. Pensando nestes fatores a área de Sistemas de Informação pode contribuir de forma satisfatória. Dados existentes em um sistema legado que já estava inoperante, foram importados em uma nova solução tecnológica, fazendo com que todo o trabalho anterior do Instituto não fosse perdido. Com isso velhos e novos documentos arquivísticos se tornam públicos e possíveis de consulta em um único sistema totalmente feito e preparado para o ambiente *web*.

5. Conclusão

O processo de migração de dados exige que não existam perdas de informações, ainda mais quando neste processo é necessária uma transformação dos dados. Devido esta exigência foi necessário um cuidado redobrado, através de pesquisas realizadas no banco de dados MySQL, foi possível criar *selects* necessários para exportar esses dados mantendo a integridade dos mesmos.

O processo de migração ocorreu de forma satisfatória, os dados foram tratados e importados com sucesso no novo ambiente AtoM. Após isso, uma reunião foi marcada com as partes interessadas, foram apresentados os resultados e algumas ressalvas. A principal é que apesar de conseguir importar com sucesso todos os dados do sistema legado no novo ambiente, ainda é necessária uma intervenção humana de um usuário. Isto se deve ao fato das informações do sistema legado estarem incompletas, o novo ambiente AtoM, exige muito mais obrigatoriedades para que as publicações sejam gravadas e fiquem disponíveis para consulta.

Como sugestão para um trabalho futuro, sugiro um melhor estudo e entendimento da ferramenta AtoM, já que o sistema legado Gallery está completamente desvendado. Focando somente no AtoM será possível entender melhor a ferramenta, tratando melhor

as informações existentes e agilizar este processo manual do usuário. Os dados importados devem ser colocados em seus devidos campos, porém não muda o fato de existirem muitos dados faltantes, que não existiam no sistema legado e que agora são de preenchimentos obrigatório no sistema AtoM.

Referências

- ALVES, R. C. V. et.al. Ciência da Informação, Ciência da Computação e Recuperação da Informação: algumas considerações sobre os métodos e tecnologias da informação utilizados ao longo do tempo. Revista Eletrônica Informação e Cognição, v.6, n.1, p.28-40, 2007. Disponível em: < https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rachel.pdf>. Acesso em: 17 out. 2017.
- BRASIL. Conselho Nacional de Arquivos. Dicionário brasileiro de terminologia arquivística. Brasília: Governo Federal, 2005. Disponível em: < http://www.arquivonacional.gov.br/images/pdf/Dicion_Term_Arquiv.pdf >. Acesso em: 17 out. 2017.
- CONARQ - Conselho Nacional de Arquivos. Coletânea da Legislação Arquivística Brasileira e Correlata. Disponível em: <http://www.conarq.gov.br/images/coletanea/jan_2016/CONARQ_legarquivos_janeiro_2016_word.pdf> Acesso em: 17 out. 2017.
- DELINGER, Giovani; Migração entre bases de dados heterogêneas: um estudo de caso em um instituto de história cultura. 2016. 12 p. Artigo (Bacharel em Sistemas de Informação) - Universidade de Caxias do Sul, Bento Gonçalves, 2016.
- DHION, C. H.; FLORES, D. Análise e aplicação do software livre ICA AtoM como ferramenta para Descrição e Acesso às informações do Patrimônio Documental e Histórico do Município de Santa Maria. 2013. Disponível em: <<http://dx.doi.org/10.18377/2316-7300/informacaoarquivistica.v3n1p24-41>>. Acesso em: 20 out. 2017.
- ELMASRI, Ramez; SHAMKANT, Navathe B. Sistemas de Banco de Dados - 6ª Ed., 2010.
- FELIPPI, Ariel. Tecnologia da informação aplicada no gerenciamento da informação histórico cultural da Serra Gaúcha. 2016. 17 p. Artigo (Bacharel em Sistemas de Informação) - Universidade de Caxias do Sul, Bento Gonçalves, 2016.
- FOWLER, Martin. Refatoração: Aperfeiçoando O Projeto de Código Existente. [S.l.]: Bookman, 2004.
- Harris, L. IPM 11g Migration Best Practices. Nexus'10. 2010.
- HELENO R. D. M., Manoel. Metodologia de Migração de Dados em um contexto de Migração de Sistemas Legados. 2009.
- KIMBALL, R., CASERTA, J. The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and. Indianapolis: Wiley Publishing, Inc. 2004.
- KIMBALL, Ralf, ROSS, Margy. O. Data Warehouse Toolkit: O Guia Definitivo de Modelagem Dimensional. New York: John Wiley & Sons, Inc, 2013.

- Kozielski, S., WREMBEL, R. (2008). *New Trends in Data Warehousing and Data Analysis*. New York: Springer Science+Business Media.
- LOPES, L. C. *A nova arquivística na modernização administrativa*. Brasília: Projeto Editorial 2009.
- MISTRY, D. (21 de Novembro de 2014). *Big Bang vs Phased – the ideal migration approach*. Disponível em: <<http://www.xceedgroup.com/xceed-blog/big-bang-vs-phased-the-ideal-migration-approach>>. Acesso em: 20 out. 2017.
- PAIVA, L.; SANTANDER, V. *Uma proposta de evolução em Sistemas Legados*, In: *Anais do WER04 - Workshop em Engenharia de Requisitos*, Tandil, Argentina, p. 201-213. 2004.
- PAPAZOGLU, Michael; HEUVEL, Willllem. *Service-Oriented Design and Development Methodology*. *International Journal of Web Engineering and Technology*, 2006.
- PINTO, H. L. M., BRAGA, J. L. *Sistemas Legados e as Novas Tecnologias: técnicas de integração e estudo de caso*, 2005, Artigo. *Revista iP – Informática Publica*, Prodabel. Disponível em: <http://www.ip.pbh.gov.br/ANO7_N1_PDF/IP7N1_mendespinto.pdf>. Acesso em: 19 out. 2017.
- PRESSMAN, Roger S.; MAXIM, Bruce R. *Engenharia de Software - Uma Abordagem Profissional - 8ª Ed*. Porto Alegre: AMGH, 2016.
- RAMALHO, José Carlos; FERREIRA, Miguel; FARIA, Luís. *Boas Práticas na Migração de Repositórios: lições aprendidas com o CALM e o ARQBASE*. In: *11º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas: “Integração, Acesso e Valor Social”*. Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas (APBAD), 2012.
- RONDINELLI, Rosely Curi. *Carta para a preservação do patrimônio arquivístico digital*. Brasília: Governo Federal, 2005. P. 1-5. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/carta.pdf>>. Acesso em: 17 out. 2017.
- SOMMERVILLE, Ian. *Engenharia de Software*. 9ª edição, São Paulo: Pearson Prentice Hall, 2016.
- SOUZA, Rodrigo Álvares; ARAKAKI, Reginaldo. *Um processo de transformação de Arquiteturas de Sistemas Legados Baseados em Reengenharia*. 2004 POLI-USP.
- STAIR, R. M.; REYNOLDS, G. W. *Princípios de Sistemas de Informação*. Editora Cengage Learning, 2005.
- STONEBRAKER, Michael. *My Top 10 Assertions About Data Warehouses*. *Communications of the ACM*, New York, v. 54, nº 5, p. 10-11, ma 2011.