

INOVAÇÃO EM PROCESSO DE ANÁLISE E SELEÇÃO DE ACERVOS DIGITAIS

José Osvaldo De Sordi

Doutor em Administração de Empresas pela Fundação Getúlio Vargas – FGV
Professor da Universidade Nove de Julho – UNINOVE
osdesordi@gmail.com (Brasil)

Gilberto Russo Jenuino

Mestre Profissional em Administração das Micro e Pequenas Empresas pela Faculdade Campo Limpo Paulista – FACCAMP
jenuino@gmail.com (Brasil)

Manuel Antônio Meireles da Costa

Doutor em Engenharia de Produção pela Universidade de São Paulo – USP
Professor na Faculdade Campo Limpo Paulista – FACCAMP
profmeireles@uol.com.br (Brasil)

RESUMO

A computação impregnada em objetos, máquinas, instalações (*pervasive computing*) é uma das características da sociedade da informação, que proporciona maior diversidade e disponibilidade na oferta de acervos de dados e informações. Atualmente o problema informacional não é mais ausência de dados e informações, mas o seu excesso, o que torna complexo e moroso o processo de análise da qualidade e seleção de conteúdos digitais, estejam estes em banco de dados ou repositório de informações. Com o objetivo de aprimorar este processo, este trabalho propõe um artefato para análise da qualidade de acervos digitais, inovador por direcionar as ações por prioridades e por contemplar especificidades tanto de bancos de dados quanto de repositórios de informações. A pesquisa é qualitativa e aplicou a técnica de grupo de discussão, percorrendo as etapas iniciais da estratégia de pesquisa *design science research* (DSR) associadas à concepção e averiguação da funcionalidade (*affordance*) de um artefato de natureza científica. Treinaram-se cinco pesquisadores na utilização do artefato, que o utilizaram para análise de cinco acervos associados a micro e pequenas empresas brasileiras, sendo: três de dados, disponíveis em banco de dados; e dois de informações, disponíveis em repositórios de informação. O artefato proposto foi avaliado por meio de uma escala de opiniões e atitudes. Os processos de pesquisa realizados colaboraram com a evolução do artefato, com a identificação e execução de pontos de melhoria que o tornaram mais funcional segundo seus usuários, adequando-o para etapa seguinte da DSR, que implica no teste de utilidade.

Palavras-chave: Repositório de informação; Banco de dado; Qualidade da informação, Qualidade dos dados.

1. INTRODUÇÃO

A computação impregnada em objetos, máquinas, instalações e na sociedade como um todo (*pervasive computing*), é uma das características da sociedade da informação, que proporcionou maior diversidade e disponibilidade na oferta de acervos de dados e informações (Gorey & Dobat, 1996). Atualmente o problema administrativo não está mais na ausência de informações, mas no seu excesso, fenômeno conhecido como *information explosion* (MA et al., 2010) ou *information overload* (Bawden & Robinson, 2009; Eppler & Megis, 2004). O excesso de informações acarreta diversos problemas ao ambiente organizacional: a) dificuldade de selecionar informação de qualidade, tornando improdutivo o ambiente de trabalho (Bawden & Robinson, 2009); b) perda de controle sobre a informação, considerando que o indivíduo tem capacidade limitada para processar e gerenciar informações (Bettman, 1979; Lazarte, 2000); e c) tomada de decisões sem contemplar a análise de todo o amplo conjunto de dados disponíveis e importantes para o processo, por questões de demanda e tempo (GAO et al., 2010).

No contexto organizacional, a título de exemplo dessa demanda, pode se imaginar um gestor da área financeira, decidindo dentre diversos serviços de provimento de informações para análise de crédito pessoal. No segmento de educação e pesquisas, considere-se, por exemplo, um gestor da área de pesquisa científica, decidindo sobre qual provedor de conteúdo contratar; na área de gestão, a escolha entre os repositórios de documentos científicos: ProQuest, EBSCO ou JStor. Em suma, trata-se de uma dificuldade crescente, independente do segmento de negócio, porte de empresa ou da área de especialização do trabalhador da informação, a qual se deseja analisar e selecionar o acervo mais adequado. Observando a literatura científica em apoio a esta demanda, encontram-se dois tipos de textos: que descrevem dimensões a serem consideradas para análises, algumas específicas para dados (Loshin, 2011; Batini et al., 2009) e outras para informações (Huang et al., 1999), de forma bastante abrangentes e genéricas; e textos voltados para análise em profundidade de uma dimensão específica, para um contexto também específico, por exemplo, um método para análise da dimensão credibilidade voltada especificamente para mídias jornalísticas (INTERNATIONAL, 2007). Na perspectiva mais pragmática, estes textos são pouco práticos e pertinentes com a demanda de análise e seleção dentre os acervos disponíveis.

Com o propósito de apoiar o processo de análise e seleção de acervos, esta pesquisa objetiva propor um artefato funcional para análise de qualidade de acervos digitais. Em função do artefato não demandar o domínio de conceitos e técnicas específicas, por exemplo, como os obtidos no exercício da

função de administrador de banco de dados ou na formação como cientista da informação, o artefato permite atender a uma comunidade de usuários (público-alvo) bastante abrangente. O público usuário do artefato pode ser generalizado como trabalhador da informação, ou seja, o profissional que se utiliza de dados e informação como parte da execução de suas atividades, sendo esse processo de utilização da informação influenciado por aspectos cognitivos (Rybckynski, 2007). Assim, o objeto desta pesquisa é o artefato proposto, sendo a importância crescente do processo de análise e seleção de acervos a justificativa técnica e social do objeto e, por consequência, da pesquisa. Este processo também delimita e especifica o contexto para o teste de funcionalidade (*affordance*) do artefato.

As diferentes tipologias de inovação sempre abrangearam entre as principais e mais citadas formas, a entidade processo. Tidd et al. (2008) descreveram as seguintes formas de inovação: (i) de produto, quando ocorre mudança nos produtos e serviços que uma empresa oferece; (ii) em processo, quando se verifica mudança na forma como os produtos ou serviços são produzidos e entregues; (iii) de posição, quando ocorre mudanças no contexto em que os produtos ou serviços são introduzidos; e (iv) de paradigma, quando são observadas mudanças nos modelos mentais subjacentes que orientam o que a empresa faz. De forma similar, Niosi et al. (1993) indicaram: novos e melhores produtos e processos, novas formas organizacionais, aplicação da tecnologia existente em novos campos, a descoberta de novos recursos e a abertura de novos mercados.

A inovação em processos, segundo Davenport (1993), é mais que a racionalização ou a simplificação, sendo importante para a redução de custos, aumento da produtividade, aumento da qualidade, redução do tempo de fabricação e de erros, ou seja, aspectos que tornam a empresa mais competitiva. Exemplo clássico de inovação de processo com geração de grande vantagem competitiva é o de Henry Ford, que em 1909 criou a linha de montagem de automóveis, revolucionando o conceito de produção industrial existente até aquele momento. De acordo com Dosi (1988) a inovação é a busca, descoberta, experimentação, desenvolvimento, imitação e adoção de novos produtos, novos processos e novas técnicas organizacionais.

Além do amplo escopo de sua aplicação, abrangendo tanto conteúdos de bases de dados quanto de repositórios de informações, o artefato proposto inova ao especificar e dimensionar as diversas dimensões a serem consideradas para análise e seleção entre acervos, ordenando-as em fases de análises, conforme as prioridades específicas do processo de análise e seleção de acervos. Desta forma o artefato abordado nesta pesquisa, conforme as definições de Damanpour e Gopalakrishnan (1991), inova na medida em que proporciona a geração, desenvolvimento e implantação de novas ideias e comportamentos (Damanpour E Gopalakrishnan, 1991).

Quanto ao tipo da inovação o artefato proposto exprime uma inovação de processo, tendo em conta que o processo de análise e seleção de acervo é significativamente aprimorado. Inovações de

processo de acordo com Sundbo e Gallouj (2000) estão relacionadas a modificações de procedimentos prescritos para a elaboração ou produção de um serviço, o que é o presente caso. Quanto ao grau de inovação pode-se considerar que o artefato exprime uma inovação incremental ao processo de análise e seleção de acervo que, segundo Leifer et al. (2003), este tipo de inovação possibilita a empresa tornar-se mais eficiente e reduzir gastos. Quanto à classe da motivação, pode-se considerar que é ‘por necessidade’, decorrente de atender ao problema gerado pela *information explosion* ou *information overload*, e ‘por mudanças demográficas’, considerando a quantidade de transações digitais decorrentes do evento do *pervasive computing*. Tanto uma justificativa quanto a outra produzem resultados benéficos: Roper e Love (2002) afirmam que a inovação motivada por necessidade não deixa de prover vantagem competitiva e Shane (2003) argumenta que mudanças demográficas são fontes de oportunidade para inovações.

No próximo capítulo, apresentam-se os conceitos necessários para o entendimento do artefato; no terceiro capítulo, descreve-se o método de pesquisa aplicado; no quarto capítulo, a descrição do artefato; no quinto capítulo, os resultados alcançados com a aplicação do artefato para análise de acervos, mais especificamente, o parecer de praticantes quanto à funcionalidade do artefato em termos de comprehensibilidade e exequibilidade das questões constituintes do artefato; e no sexto capítulo, as conclusões.

2. CONCEITOS E ABSTRAÇÕES LÓGICAS ASSOCIADAS AO ARTEFATO

No campo da epistemologia da ciência, o termo acervo configura uma coleção de dados e/ou informações organizadas segundo critérios de interesse, podendo ser públicos ou privados (Lino, 2005). Ao longo do texto utiliza-se o termo acervo tanto para caracterizar um conjunto de relatórios digitais ou de bases de dados digitais. O Artefato proposto nesta pesquisa volta-se para qualidade de acervos, demandando o uso de termos e conceitos associados às entidades: tecnologia para guarda de dado, tecnologia para guarda de informação e qualidade dos acervos. Assim, nesta seção, definem-se: a) as dimensões para análise da qualidade, igualmente aplicadas tanto a dados quanto a informações; b) para dados, define-se a tecnologia do sistema gerenciador de banco de dados e as bases de dados; c) para informações, conceitua-se a tecnologia de repositório de informações.

A literatura acadêmica não apresenta um consenso quanto às dimensões de qualidade da informação que devem ser consideradas. As variações ocorrem em função da taxonomia adotada dentre as muitas disponíveis na literatura sobre o tema. Garvin (1988), Salmela (1997), Tozer (1999) e

Huang et al. (1999) são alguns dos autores que esquematizaram diferentes conjuntos de dimensões. Huang et al. (1999), por exemplo, desenvolveram quatro categorias de dimensões: (1) qualidade intrínseca: acurácia, objetividade, credibilidade, e reputação; (2) qualidade de acessibilidade: acesso e segurança; (3) qualidade contextual: relevância, valor-agregado, economia de tempo, completeza e quantidade de dados; e (4) qualidade representacional: interpretabilidade, facilidade de uso, representação concisa e representação consistente. Para a análise da qualidade de dados observa-se a mesma dificuldade: muita diversidade de dimensões e abordagens (Batini *et al.*, 2009).

Para disponibilização de informações, observa-se a proliferação da tecnologia de repositórios institucionais. O repositório institucional (RI) é um sistema computacional que auxilia as organizações a colecionar, administrar, distribuir, e conservar seus ativos digitais (Smith, 2002). Os RI empregam o conceito de interoperabilidade denominado *Open Archives Initiative - Protocol for Metadata Harvesting* (OAI-PMH), que tem como características: estruturas de dados codificadas em XML, transmitidas em HTML e com metadados sobre os conteúdos da pesquisa no formato Dublin Core. Entre os benefícios da adoção destes novos padrões está a facilidade de indexação, divulgação e localização dos conteúdos disponíveis nos RI. O conteúdo do RI é caracterizado por informação, normalmente expressa na forma de documentos digitais, que denominaremos ao longo desta pesquisa de relatório.

Para disponibilização de dados utiliza-se o sistema gerenciador de banco de dados (SGBD), definido como um sistema computadorizado para manutenção e registro de dados (DATE, 2004). O SGDB permite a criação e gestão de diferentes bancos de dados, que são constituídos por bases de dados. A base de dados é um conjunto de dados inter-relacionados, organizados de forma a permitir a recuperação de um conjunto de evidências associados a um evento ou transação específica (Cunha, 1989).

Associadas ao campo da *design school*, esta pesquisa utiliza-se da estratégia de pesquisa *design science research* (DSR), descrita no próximo capítulo de “método da pesquisa”, bem como do conceito de *affordance*. O conceito de *affordance* aborda as dualidades da relação entre sujeito e objeto. *Affordance* está diretamente associado aos fatos decorridos da ação, não se trata de percepção subjetiva, mas da interação objetiva e direta do sujeito junto ao artefato (Gaver, 1996). De forma pragmática, *affordance* indica a praticidade, a funcionalidade do objeto para o seu propósito, segundo seus usuários. Serve como meio para análise e previsão da intensidade de uso do artefato (Kuo *et al.*, 2013).

3. MÉTODO DA PESQUISA

Tipo de Pesquisa. A pesquisa é qualitativa com base em dados primários, oriundos das respostas de cinco usuários que validaram o artefato proposto. Foi utilizada a técnica de grupo de discussão, conforme procedimentos de Canales e Peinado (1994). Objetivando uma pesquisa mais rigorosa, em termos de concepção e validação do artefato, submeteu-se o artefato ao tratamento que preceitua a abordagem *design science research*. Segundo Hevner, March, Park e Ram (2004, p.81) este tipo de pesquisa “aborda problemas importantes, ainda não resolvidos, de forma única e inovadora, ou problemas resolvidos, porém de forma mais eficaz ou eficiente”.

Amostras. Para fins de teste de funcionalidade (*affordance*) e apresentação detalhada das partes constituintes do artefato, aplicou-se o mesmo para análise de: três acervos de dados e dois acervos de informações, todos voltados às micro e pequenas empresas (MPE) brasileiras. Os cinco acervos estão descritos no Apêndice A. A temática dos acervos foi em função dos interesses específicos dos trabalhadores da informação que testaram a funcionalidade do artefato, no caso, pesquisadores de um grupo de pesquisa cujo interesse central é a gestão das MPE.

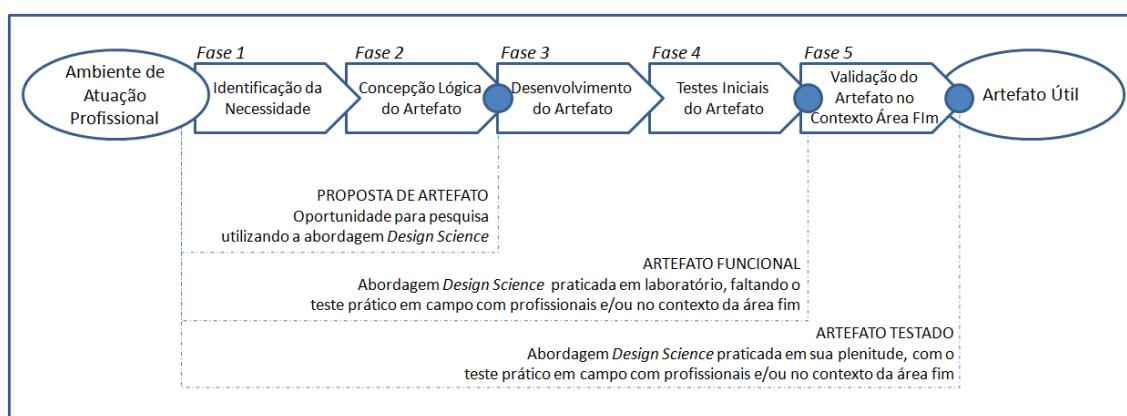
Procedimentos para construção do artefato. O artefato para análise da qualidade de conteúdos digitais, armazenados em banco de dados ou em repositórios de informações, foi desenvolvido a partir da coleta, análise, interpretação e associação de dois conjuntos de conceitos e informações: a) métodos para análise de qualidade de bancos de dados e b) dimensões para análise de qualidade da informação. O primeiro conjunto é de natureza mais pragmática, proveniente de pesquisadores e praticantes da área da Ciência da Computação e da Administração (de sistemas de informação). O segundo conjunto é de natureza mais conceitual, proveniente predominantemente da área de Ciência da Informação. Após a leitura intensiva e análise detalhada destes conteúdos, identificaram-se as dimensões de qualidade mais pertinentes com a proposta do artefato, ou seja, as aplicáveis e relevantes tanto para caracterização da qualidade de dados quanto de informações. As treze dimensões de qualidade abordadas pelo artefato são: contextualização, abrangência, acessibilidade, confidencialidade, precisão, acurácia, confiabilidade, atualidade, ineditismo, originalidade, existência, identidade e integridade.

Observando o comportamento de trabalhadores da informação que procuram por acervos de interesse, observou-se que estes praticam diferentes momentos de pesquisa, com diferentes interesses.

Tomando como exemplo as fases de pesquisa por textos, praticadas por pesquisadores, pode se estabelecer uma analogia como as fases iniciais de leitura de seleção de texto do tipo *scanning*, *skimming* que precedem as leituras do tipo intensiva e reflexiva (De Sordi, 2013). No primeiro momento da pesquisa, o trabalhador da informação objetiva considerar a pertinência do acervo em relação ao tema que se busca para, posteriormente, realizar uma análise efetiva da qualidade dos dados ou informações do acervo. Por essa razão, o artefato foi concebido com dois estágios de aplicação, composto por dois conjuntos de questões: primeiro, composto por questões voltadas para análise do contexto do acervo, se pertinente como o tema que se deseja; segundo, contempla questões específica para análise de qualidade dos dados ou informações do acervo.

É importante observar que há diferentes alternativas para teste da construção e validação da lógica de artefatos, por exemplo, a encontrada para proposição de uma escala, que deve estar associada a uma variável pertinente a um problema ou oportunidade do ambiente de negócios (Iivari, 2007; Kuechler & Vaishnavi, 2008). A figura 1 descreve os diferentes momentos do ciclo de desenvolvimento do artefato; os testes para construção e validação do artefato estão declarados como “artefato funcional”, enquanto que os testes para análise de funcionalidade junto aos profissionais “praticantes” estão descritos como “artefato testado”.

Figura 1 - Fases do ciclo de desenvolvimento do artefato



O artefato, na forma de questões associadas às dimensões de qualidade, com a descrição de procedimentos para análise ou mensuração dos tópicos vinculados às questões, foi inicialmente testado pelos próprios proponentes. Posteriormente ao “teste de laboratório”, conforme descrito na fase 4 da Figura 1, realizou-se mais um teste em campo do artefato, ou seja, a descrita na fase 5. Para o teste prático do artefato em campo, os pesquisadores integrantes de um grupo de pesquisa com interesse na gestão das MPE utilizaram o artefato para análise de cinco acervos pertinentes às MPE brasileiras,

sendo: três bancos de dados e dois repositórios de informações. Este teste proporcionou diversas correções e aprimoramentos, em especial na forma de apresentar as questões e os procedimentos para análise e mensuração dos aspectos característicos de cada dimensão em análise, considerando tanto o contexto dos bancos de dados quanto dos repositórios de informações.

Procedimentos para validação do artefato. A verificação da funcionalidade do artefato seguiu as diretrizes apontadas por Hevner et al., 2004. O ‘teste de campo’ para validação do artefato foi realizado por uma equipe de cinco pesquisadores integrantes de um grupo de pesquisas da área de Administração, com interesse na gestão das MPE. Em novembro de 2013, os desenvolvedores apresentaram e disponibilizaram o artefato para esses cinco pesquisadores e indicaram cinco acervos sobre MPE a serem analisados. Cada um dos cinco pesquisadores do grupo de pesquisa se comprometeu em analisar um dos acervos com o a utilização do artefato e, posteriormente, preencher um formulário relatando a experiência com a utilização do artefato.

Variáveis. As variáveis coletadas referem-se à opinião dos cinco pesquisadores que desempenharam a função de avaliadores do artefato a partir da experiência destes no ‘teste de campo’ para análise de um acervo sobre MPE. As variáveis analisadas contemplam o parecer dos avaliadores, expressos por meio de uma escala Likert (1-5), para cada uma das treze dimensões de qualidade consideradas: acessibilidade, confidencialidade, precisão, acurácia, confiabilidade, atualidade, ineditismo, originalidade, existência, identidade e integridade.

Tratamento dos Dados. Os dados receberam tratamento qualitativo por meio de técnicas não-paramétricas. Para avaliar o grau de aderência (GA) dos respondentes a cada proposição e ao conjunto total das proposições, foi utilizado um oscilador que mostra a força relativa entre a discordância ((respostas Muito difícil (MD) + respostas Difícil (D) + respostas Indiferente) / 2) e a concordância ((respostas Muito fácil (MF) + respostas Fácil (F) + respostas Indiferente (I)) / 2). O oscilador segue a proposta de Wilder (1981) e gera um número entre um mínimo de zero e um máximo de 100. Níveis acima de 50 indicam tendência concordante e, abaixo, tendência discordante. Níveis iguais ou acima de 80 indicam alto nível de aderência e níveis iguais ou abaixo de 20 indicam falta de aderência. A fórmula geral é:

$$GA = 100 - \left(\frac{100}{\frac{MF + F + \frac{1}{2}I}{MD + D + \frac{1}{2}I} + 1} \right)$$

Limitações do Método e Dificuldades deste Estudo. Nesta pesquisa realizaram-se os “testes iniciais do artefato” dentro da abordagem *design science research*, conforme indicado na Figura 1. Os cinco pesquisadores envolvidos nos testes não caracterizam plenamente uma “validação do artefato no contexto da área fim”, conforme indicado na fase 5, da Figura 1. Para que o teste ocorrido fosse considerado como um teste de campo completo, os cinco pesquisadores deveriam ter testado todas as questões do artefato, inclusive as cinco iniciais para pré-análise dos acervos, abrangendo as dimensões contextualização e abrangência. Isso não ocorreu em função da pouca disponibilidade de acervos sobre MPE brasileiras. Para assegurar maior envolvimento dos cinco pesquisadores no processo de testes do artefato, optou-se por indicar cinco acervos conhecidos sobre as MPE brasileiras, quando da apresentação do artefato, e a solicitação de utilização do artefato pelos pesquisadores.

3.1 O Artefato

O Quadro 1 apresenta a versão final do artefato contendo dois grupos de questões, aplicado em dois momentos distintos. O primeiro grupo, aplicado na fase de pré-análise do acervo, contendo questões que buscam compreensão da pertinência do acervo com o tema que se deseja obter conteúdo por intermédio do acervo; em havendo pertinência, aplica-se o segundo grupo de questões, voltado à análise técnica e detalhada do acervo. As questões presentes no Quadro 1 são as resultantes da experiência realizada, ou seja, de: a) proposição de questões norteadoras; b) testes destas por pesquisadores de um grupo de pesquisa com interesse na análise de acervos; e c) aprimoramento do artefato a partir da experiência e do *feedback* dos pesquisadores que utilizaram o artefato inicialmente proposto.

Quadro 1 - Artefato para análise e seleção de acervos de dados ou informações

Fase 1 - Pré-análise	1. Dimensão Contextualização do conteúdo
	1 1.1 O contexto do tema abordado no acervo é o mesmo do tema que se procura?
	2 Observação: definir questão(ões) que permita(m) esta análise. O pesquisador deverá desenvolver quantas questões forem necessárias para contextualizar o tema de interesse, que se deseja encontrar, no conteúdo dos acervos a serem analisados. O tema pode ser mais amplo ou mais específico, por exemplo, “logística” e “logística reversa”, dependendo exclusivamente do interesse do pesquisador e da capacidade deste em desenvolver questões que permitam o nível de detalhamento desejado
	2. Dimensão Abrangência / Escopo do conteúdo
	3 2.1 O conteúdo abrange a diversidade de assuntos sobre o tema que se deseja, ou seja, nem muito assuntos a mais, nem muitos assuntos a menos? (cobertura de assuntos – pesquisa horizontal)
	4 Observação: definir questão(ões) que permita(m) esta análise. Desenvolver questões horizontais para análise do tema de interesse, ou seja, assuntos ou vértices do conteúdo. O pesquisador poderá desenvolver quantas questões achar necessário para analisar a abrangência da dimensão

Fase 2 - Análise Detalhada do Acervo	<p>3. Dimensão Acessibilidade do conteúdo</p> <p>5 3.1 Esforços necessários para se acessar o conteúdo?</p> <p>Refere-se ao número de cliques do mouse. Conta-se da página inicial do site até a apresentação da tela de seleção de pesquisa do banco de dados ou até a o link do relatório desejado no repositório.</p> <p>1 3.2 Qual o tempo necessário para acessar o conteúdo?</p> <p>Tempo médio para três acessos realizados em momentos distintos, com registro de minutos e segundos. Conta-se o tempo a partir do último clique, ou seja, tempo para processamento da busca (<i>query</i>) no banco de dados ou do <i>download</i> do relatório a partir do repositório</p> <p>2 3.3 Qual a disponibilidade do serviço de provisionamento de conteúdo?</p> <p>Averigua o quanto o site está disponível aos usuários, pela fórmula: Total de acessos realizados (conexões realizadas) / Total de acessos solicitados (conexões realizadas + conexões não realizadas).</p> <p>4. Dimensão Confidencialidade / Privacidade do conteúdo</p> <p>3 4.1 Há necessidade de identificação do pesquisador para o acesso ao repositório ou base de dados?</p> <p>Sim (“S”), se requer identificação do usuário para acesso ao conteúdo; ou Não (“N”), se não requer identificação do usuário para acesso.</p> <p>1 4.2 Há alguma limitação no retorno da pesquisa (quantidade máxima de dados ou páginas)?</p> <p>Sim (“S”), se apresenta limitação para pesquisa de dados (quantidade de registros apresentados) ou de relatório (quantidade de páginas); ou Não (“N”), se não apresenta limitação para pesquisa de dados (quantidade de registros apresentados) ou de relatório (quantidade de páginas).</p> <p>5. Dimensão Precisão do conteúdo</p> <p>4 5.1 Os conteúdos estão no nível de detalhamento desejado, ou seja, nem mais, nem menos? (cobertura do detalhamento necessário – pesquisa vertical)</p> <p>Sim (“S”), se o fato elementar descrito no banco de dados ou no conteúdo do relatório está no nível de detalhamento que se necessita; ou Não (“N”), se o conteúdo está, exclusivamente, detalhado demais ou genérico demais.</p> <p>5.2 Permite ao pesquisador acessar estes valores unitariamente ou por faixas intervalares?</p> <p>Para o banco de dados: Sim (“S”), se permite consulta por faixa intervalar, por exemplo, por quantidade de funcionário e ou por faturamento da empresa; ou Não (“N”), se não há disponibilidade desta função/recurso. Para o repositório: Sim (“S”), se permite visualização das seções e navegação rápida no relatório, por exemplo, um sumário com navegação automática ao clicar o número da página ou título da estrutura; ou Não (“N”), se não há disponibilidade desta função/recurso.</p> <p>6. Dimensão Acurácia / Veracidade do conteúdo</p> <p>5 6.1 O método declarado indica o período da coleta / geração da análise?</p> <p>6 Sim (“S”), se há a declaração da data ou período em que os dados foram coletados/colhidos ou a informação foi gerada; ou Não (“N”), se não há esta informação disponível no acervo.</p> <p>7 6.2 O método declarado indica a fonte da coleta dos insumos?</p> <p>8 Sim (“S”), se há a declaração da fonte para a coleta dos dados ou para a geração da informação; ou Não (“N”), se não há esta informação disponível no acervo.</p> <p>9 6.3 O método declarado indica o instrumento utilizado na coleta / para análise?</p> <p>10 Sim (“S”), se há a declaração da metodologia ou técnica utilizada para coleta dos dados ou da metodologia empregada para geração da informação; ou Não (“N”), se não há esta informação disponível no acervo</p> <p>7. Dimensão Confiabilidade do conteúdo</p> <p>11 7.1 Qual a percepção da sociedade e/ou academia científica com relação à fonte geradora do conteúdo?</p> <p>12 Utilizando softwares de busca na Internet para verificar se a fonte geradora do conteúdo está associada a eventos ou fatos que possam denegrir a confiabilidade na mesma. Sim (“S”), para nada encontrado que possa comprometer a confiabilidade na fonte geradora do conteúdo; Não (“N”), para situações encontradas que possam comprometer a confiabilidade na fonte geradora do</p>
---	--

	<p>conteúdo.</p> <p>13 7.2 Qual a percepção da sociedade e/ou academia científica com relação à fonte que disponibiliza o conteúdo (caso distintas, uma gera e outra publica)?</p> <p>14 Utilizando softwares de busca na Internet para verificar se a fonte que disponibiliza o conteúdo está associada a eventos ou fatos que possam denegrir a confiabilidade na mesma. Sim (“S”), para nada encontrado que possa comprometer a confiabilidade na fonte que disponibiliza o conteúdo; Não (“N”), para situações encontradas que possam comprometer a confiabilidade na fonte que disponibiliza o conteúdo</p>
	<p>8. Dimensão Atualidade / Temporalidade do conteúdo</p> <p>1 8.1 Qual é o tempo médio (período) de atualização do conteúdo referente ao acervo em análise? Analisar a atualização do banco de dados ou do repositório em termos de conteúdo, por exemplo: conteúdo gerado/coletado/colhido diariamente, mensalmente, trimestralmente, semestralmente, anualmente, e etc.</p> <p>1 8.2 Qual o período de tempo decorrido entre a geração e a publicação do conteúdo? Tempo médio decorrido entre a geração/coleta do conteúdo e a publicação efetiva do conteúdo, ou seja, disponível para consulta ou download. Exemplos: atraso de um dia, atraso de um ano, concomitante com a geração/coleta do conteúdo</p>
	<p>9. Dimensão Ineditismo / Raridade do conteúdo</p> <p>2 9.1 Estes conteúdos estão disponíveis em outros acervos (semelhantes e não cópias)? Realizar busca na Internet com relação aos termos associados ao título do relatório ou ao tema do banco de dados. Sim (“S”), para presença de conteúdo semelhante; ou Não (“N”), quando não for encontrado conteúdo semelhante.</p>
	<p>10. Dimensão Originalidade do conteúdo</p> <p>1 10.1 Os valores registrados no banco de dados ou no repositório de informação foram obtidos junto à fonte primária? Analisar se o conteúdo do banco de dados ou do repositório é de fonte primária ou secundária. Entende-se por fonte primária os dados, armazenados no banco de dados, ou o conteúdo utilizado para geração da informação, serem provenientes da fonte mais direta possível em relação ao objeto tema do acervo. Assinalar (“S”) para fonte primária e (“N”) para fonte secundária.</p>
	<p>11. Dimensão Existência do conteúdo</p> <p>1 11.1 Este conteúdo virtual existe em várias localidades (cópias)? Analisar se o conteúdo do acervo possui cópias físicas ou virtuais. Sim (“S”), para presença de cópias; ou Não (“N”), quando não houver cópias.</p>
	<p>12. Dimensão Identidade do conteúdo</p> <p>1 12.1 Há metadados/metainformações disponíveis que auxiliem na compreensão e identificação do dado ou informação? Sim (“S”), se há a declaração de metadados e/ou metainformações para o acervo; ou Não (“N”), se não há disponibilidade de metadados e/ou metainformações para o acervo.</p>
	<p>13. Dimensão Integridade do conteúdo</p> <p>1 13.1 Percebe-se o registro de dados indevido nos campos/colunas, no caso dos bancos de dados, ou de páginas, no caso de relatórios?</p> <p>2 Analisar se há erros ou omissão nos documentos do repositório, exemplo, falta de páginas, frases propositalmente truncadas / excluídas; para o banco de dados se há registro de dados incompatível com o campo ou ainda se há falta de registro (campos vazios/nulos) entre outros. Sim (“S”), se há a integridade; ou Não (“N”), quando nota-se problemas de integridade do conteúdo.</p>

Em janeiro de 2014, os desenvolvedores do artefato reuniram-se com o grupo de pesquisa com interesses nas MPE brasileiras a fim de coletar o *feedback* com relação à utilização do artefato. Este processo resultou em novas adaptações e aprimoramento do artefato, em especial na forma de redigir e apresentar as questões, bem como os procedimentos operacionais para análise ou mensuração do tópico abordado em cada questão.

4. RESULTADOS

Nos testes, os pesquisadores tinham como interesse a busca por acervos associados às MPE brasileiras. Para esta análise, definiram-se três questões para contextualização do acervo, todas descritas no Quadro 2, conjuntamente, com as respectivas justificativas dos pesquisadores para formulação de cada questão. Ainda na fase de pré-análise, o artefato solicita aos seus usuários, que definia um conjunto de questões que consiga averiguar a abrangência do tratamento do tema pelo acervo, ou seja, a diversidade das características e especificidades do tema abordado pelo acervo em comparação com as demandas do pesquisador. Para a experiência realizada, as questões de abrangência também estão descritas no Quadro 2, abrangendo questões específicas para análise horizontal de três perspectivas: temporal (décadas), territorial (Estados brasileiros) e por atividade econômica (segmentos de negócios).

Quadro 2 -Exemplos de questões específicas do artefato para fase de pré-analise do acervo

1. Dimensão Contextualização do conteúdo
1.1 Neste acervo há conteúdo referente à quantidade de funcionários da empresa? Justificativa para a questão: o grupo trabalha com duas definições de MPE, a do SEBRAE, que utiliza quantidade de funcionários das empresas, e a da Receita Federal, que utiliza faturamento anual
1.2 Neste acervo há conteúdo referente ao faturamento da empresa? Justificativa para a questão: o grupo trabalha com duas definições de MPE, a do SEBRAE, que utiliza quantidade de funcionários das empresas, e a da Receita Federal, que utiliza faturamento anual
1.3 Neste acervo há alguma outra forma distintiva que caracterize MPE? Justificativa para a questão: identificar outra forma de classificação das empresas quanto ao seu porte, que seja distinta de quantidade de funcionário ou faturamento anual
2. Dimensão Abrangência/escopo do conteúdo
2.1 O Conteúdo abrange MPEs do Brasil, de alguns estados, de alguns municípios ou outra forma de expressão de escopo territorial? Justificativa para a questão: identificar o escopo territorial das empresas descritas no acervo
2.2 O conteúdo aborda MPEs de forma geral ou as diferencia de alguma forma por segmentos de negócio? Justificativa para a questão: identificar o escopo do âmbito de negócio das empresas descritas no acervo
2.3 O conteúdo abrange séries históricas das MPE's por séculos, por décadas, por ano, por semestre, por mês, por dia ou por algum outro escopo temporal? Justificativa para questão: identificar o escopo do âmbito temporal das empresas descritas no acervo

Uma vez o acervo considerado pertinente segundo as duas dimensões de análise aplicadas na primeira fase de pré-análise (contextualização e abrangência), o mesmo era analisado segundo as

demais onze dimensões de qualidade, indicadas na parte inferior do Quadro 1, denominada de “Fase 2 - Análise detalhada do acervo”. Esta segunda fase abrange a análise das dimensões: acessibilidade, confidencialidade, precisão, acurácia, confiabilidade, atualidade, ineditismo, originalidade, existência, identidade e integridade. O Quadro 3 apresenta os resultados das análises de qualidade conduzida para os cinco acervos referente às MPE brasileiras, realizadas pelos cinco membros do grupo de pesquisa que auxiliou no teste de funcionalidade do artefato.

Quadro 3 -Resultados da aplicação do artefato para análise de cinco acervos referente às MPE brasileiras

		Acervo Analisado				
		BD-1	BD-2	BD-3	RI-1	RI-2
Fase 1 - Pré-análise	1. Dimensão Contextualização do conteúdo					
	1 1.1 Descreve quantidade de funcionários das empresas?	S	S	S	S	S
	2 1.2 Descreve faturamento das empresas?	N	N	N	N	N
	3 1.3 Há outra forma de caracterização da MPE?	N	N	N	N	N
	2. Dimensão Abrangência/escopo do conteúdo					
	2.1 Há conteúdo diferenciado por escopo territorial?	S	S	S	S	S
	2.2 Há conteúdo diferenciado por segmento de negócio?	S	S	S	S	S
Fase 2 - Análise Detalhada do Acervo	2.3 Há conteúdo diferenciado por escopo temporal?	S	S	N	S	S
	3. Dimensão Acessibilidade do conteúdo					
	3.1 Esforços necessários para se acessar o conteúdo? (quantidade de cliques com o mouse)	6	3	3	6	4
	3.2 Qualo tempo necessário para acessar o conteúdo?	27,03'	40,84'	1,92'	15,99'	3,23'
	3.3 Qual a disponibilidade do serviço de provisionamento do conteúdo?	3/4	4/4	4/4	4/4	4/4
	4. Dimensão Confidencialidade/Privacidade do conteúdo					
	4.1 Há necessidade de identificação do pesquisador para o acesso ao repositório ou base de dados?	N	S	N	N	N
	4.2Há alguma limitação no retorno da pesquisa (quantidade máxima de dados ou páginas)?	S	S	S	N	N
	5. Dimensão Precisão do conteúdo					
	5.1 Os conteúdos estão no nível de detalhamento desejado, ou seja, nem mais, nem menos?	N	N	N	N	N
	5.2 Permite ao pesquisador acessar estes valores unitariamente ou por faixas intervalares?	S	S	S	S	S
	6. Dimensão Acurácia/Veracidade do conteúdo					
	6.1 O método declarado indica o período da coleta/geração da análise?	N	S	S	N	S
	6.2 O método declarado indica a fonte dos insumos?	S	S	S	S	S
	6.3 O método declarado indica o instrumento utilizado na coleta/para análise?	S	S	S	N	S
	7. Dimensão Confiabilidade do conteúdo					
	7.1 Qual a percepção da sociedade e/ou academia científica com relação à fonte geradora do conteúdo?	S	S	S	S	S

	7.2 Qual a percepção da sociedade e/ou academia científica com relação à fonte que disponibiliza o conteúdo (caso distintas, uma gera e outra publica)?	S	S	S	S	S
	8. Dimensão Atualidade/Temporalidade do conteúdo					
	8.1 Qual o tempo médio (período) de atualização do conteúdo referente ao acervo em análise?	ANO	ANO	ANO	ANO	MÊS
	8.2 Qual o período de tempo decorrido entre a geração e a publicação do conteúdo?	2 anos	2 anos	0	1 a 2 anos	0
	9. Dimensão Ineditismo/Raridade do conteúdo					
	9.1 Estes conteúdos estão disponíveis em outros acervos (semelhantes e não cópias)?	S	S	S	S	S
	10. Dimensão Originalidade do conteúdo					
	10.1 Os valores registrados no banco de dados ou no repositório de informação foram obtidos junto à fonte primária?	N	S	S	N	S
	11. Dimensão Existência do conteúdo					
	11.1 Este conteúdo virtual existe em várias localidades (cópias)?	N	N	N	S	S
12. Dimensão Identidade do conteúdo						
12.1 Há metadados/metainformações disponíveis que auxiliem na compreensão e identificação do dado ou informação?						
13. Dimensão Integridade do conteúdo						
13.1 Percebe-se o registro de dados indevido nos campos/colunas, no caso de banco de dados, ou de páginas, no caso de relatórios?						

Legenda: S=sim, N=não; **BD-1:** Instituto Brasileiro de Geografia e Estatística (IBGE); **BD-2:** Ministério do Trabalho e Emprego (MTE); **BD-3:** Fundação Sistema Estadual de Análise de Dados (SEADE); **RI-1:**Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE); **RI-2:** Federação do Comércio de Bens, Serviços e Turismo(FECOMERCIO-MG)

Solicitou-se aos cinco membros do grupo de pesquisa da MPE, que ao utilizarem o artefato para análise dos cinco acervos, que analisassem as questões do artefato em termos da dificuldade encontrada para: a) compreensão da pergunta, b) execução dos procedimentos solicitados para fins de análise e/ou mensuração do tópico associado à dimensão da qualidade. Solicitou-se os pareceres apenas para as perguntas da fase 2, da análise detalhada do acervo, considerando-se que as perguntas da fase 1, de pré-análise do acervo, são específicas para cada temática e formuladas pelos próprios pesquisadores com interesse nos acervos. O Quadro 4 apresenta a consolidação das respostas dos cinco membros do grupo de pesquisa às 19 perguntas da fase 2.

Do total das avaliações providas pelos cinco membros do grupo de pesquisa, nenhuma foi assinalada como “muito difícil” e seis foram assinaladas como “difícil”, conforme destacado na penúltima linha do Quadro 4. Dentre as seis questões assinaladas como “difícil”, quatro foram assinaladas para questões da dimensão confiabilidade, sendo uma referente à dificuldade de compreensão e três para mensuração da confiabilidade. De forma geral, considerando todas as proposições o artefato foi considerado fácil de utilizar com aderências bem elevadas: 96,8 quanto à

compreensão e 90,0 quanto à mensuração. A proposição com baixo nível de aderência quanto à compreensão foi a proposição 7.2, referente à ‘percepção da sociedade e/ou academia científica com relação à fonte que disponibiliza o conteúdo’, com uma aderência de 70. Quanto às proposições de mensuração cabe destacar algumas com aderência inferior a 80:

- duas referentes à dimensão Confiabilidade (a proposição 7.1, referente à ‘percepção da sociedade e/ou academia científica com relação à fonte geradora do conteúdo’ e a proposição 7.2, referente à ‘percepção da sociedade e/ou academia científica com relação à fonte que disponibiliza o conteúdo’). Os valores de aderência foram, respectivamente, 70 e 50;
- uma referente à dimensão Existência (a proposição 11.1 referente ao ‘conteúdo virtual existente em várias localidades’).

Quadro 4 - Parecer dos profissionais que testaram o artefato para análise de acervos

	COMPREENSÃO					Grau de aderência	MENSURAÇÃO					Grau de aderência
	Muitodifícil	Difícil	Indiferente	Fácil	MuitoFácil		Muitodifícil	Difícil	Indiferente	Fácil	MuitoFácil	
3. Dimensão Acessibilidade												
3.1 Esforços necessários para se acessar o conteúdo?				2	3	100,0				4	1	100,0
3.2 Qual o tempo necessário para acessar o conteúdo?				2	3	100,0				4	1	100,0
3.3 Qual a disponibilidade do serviço de provisionamento de conteúdo?				4	1	100,0			1	3	1	90,0
4. Dimensão Confidencialidade												
4.1 Há necessidade de identificação do pesquisador para o acesso ao repositório ou base de dados?				4	1	100,0				3	2	100,0
4.2 Há alguma limitação no retorno da pesquisa (quantidade máxima de dados ou páginas)?				4	1	100,0				3	2	100,0
5. Dimensão Precisão												
5.1 Os conteúdos estão no nível de detalhamento desejado, ou seja, nem mais, nem menos?			1	4		90,0				5		100,0
5.2 Permite ao pesquisador acessar estes valores unitariamente ou por faixas intervalares?				4	1	100,0				4	1	100,0

6. Dimensão Acurácia												
6.1 O método declarado indica o período da coleta / geração da análise?				3	2	100,0				4	1	100,0
6.2 O método declarado indica a fonte da coleta dos insumos?				4	1	100,0			1	4		90,0
6.3 O método declarado indica o instrumento utilizado na coleta / para análise?			1	3	1	90,0			2	3		80,0
7. Dimensão Confiabilidade												
7.1 Qual a percepção da sociedade e/ou academia científica com relação à fonte geradora do conteúdo?				5		100,0		1	1	3		70,0
7.2 Qual a percepção da sociedade e/ou academia científica com relação à fonte que disponibiliza o conteúdo (caso distintas, uma gera e outra publica)?	1	1	3			70,0		2	1	2		50,0
8. Dimensão Atualidade												
8.1 Qual é o tempo médio (período) de atualização do conteúdo referente ao acervo em análise?				5		100,0				5		100,0
8.2 Qual o período de tempo decorrido entre a geração e a publicação do conteúdo?				5		100,0		1		4		80,0
9. Dimensão Ineditismo												
9.1 Estes conteúdos estão disponíveis em outros acervos (semelhantes e não cópias)?				5		100,0			1	4		90,0
10. Dimensão Originalidade												
10.1 Os valores registrados no banco de dados ou no repositório de informação foram obtidos junto à fonte primária?				5		100,0				5		100,0
11. Dimensão Existência												
11.1 Este conteúdo virtual existe em várias localidades (cópias)?				5		100,0		1	1	3		70,0
12. Dimensão Identidade												
12.1 Há metadados / metainformações disponíveis que auxiliem na compreensão e identificação do dado ou informação?				3	2	100,0				4	1	100,0
13. Dimensão Integridade												
13.1 Percebe-se o registro de dados indevido nos campos/colunas, no caso dos			1	4		90,0			1	4		90,0

bancos de dados, ou de páginas, no caso de relatórios?											
TOTAL :	0	1	4	74	16		0	5	9	71	10
Grau de Aderência (GA)	3			92		96,8	9,5			85,5	90,0

Ao indagar os respondentes sobre estes pontos de dificuldade, observou-se que a característica geradora da dificuldade foi a subjetividade do aspecto central empregado para análise da confiabilidade: “eventos ou fatos que possam denegrir a confiabilidade” da fonte geradora ou da fonte que disponibiliza o conteúdo. A dificuldade da análise está na própria natureza subjetiva da confiabilidade. Apesar da dificuldade da análise desta característica associada ao acervo, optou-se por manter a questão no artefato, em função: a) da sua importância para o não uso de dados ou informações provenientes de fontes de baixa confiabilidade; b) da sua serventia em termos de demandar a reflexão e análise, mesmo que subjetiva, de um aspecto importante da qualidade dos acervos.

5. CONCLUSÕES

As atividades desenvolvidas permitiram atender ao objetivo da pesquisa: propor um artefato funcional para análise de qualidade de acervos digitais, constituídos por dados ou por informações. Para fins de testes iniciais de funcionalidade do artefato e apresentação detalhada de suas partes constituintes, descreveu-se a aplicação do mesmo para análise de três acervos de dados e dois de informações, todos associados às MPE. Após a aplicação ampla do artefato pelos próprios desenvolvedores, os mesmos cinco acervos foram objeto de análise de um grupo de cinco pesquisadores, que também utilizaram o artefato e forneceram *feedbacks* que auxiliaram no aprimoramento do artefato.

Apesar de ser desenvolvido a partir de diferentes modelos disponíveis para análise da qualidade de dados e de informações, o artefato volta-se ao processo de análise de qualidade de uma entidade mais abrangente: o acervo de conteúdo digital, caracterizado por um conjunto de relatórios ou de bases de dados. Um primeiro nível de diferenciação, de artefato para este fim, está na ordenação de questões em momentos distintos: a fase de pré-análise do acervo, com questões que buscam compreender duas dimensões (contextualização e abrangência) associadas à pertinência do acervo com o tema que se deseja obter mais conteúdo; e as questões da fase de análise detalhada do acervo, abrangendo outras onze dimensões para decisão de seleção e uso do acervo. Esta estrutura apresenta uma ordenação bastante lógica, escalonando e evitando análises desnecessárias.

Além da ordem para análise das dimensões de qualidade, algumas das dimensões para análise da qualidade de dados ou de informações não se aplicam ao contexto das coleções ou dos acervos, por serem específicas para análise individual de um relatório ou de um banco de dados. Cita-se como exemplo, a dimensão coesão para análise da qualidade da informação (Huang et al., 1999), que requer maior envolvimento do leitor com a informação, demandando a leitura do relatório. Em suma, os acervos constituem-se em entidades distintas de seus elementos constituintes, sejam eles relatórios ou base de dados. Assim como há técnicas e fases para leitura seletiva, abrangendo as fases da leitura do tipo *scanning*, *skimming*, intensiva e reflexiva (De Sordi, 2013), que contemplam diferentes aspectos da informação, observou-se que há necessidade de considerar algumas dimensões em fases específicas para análise de acervos. O artefato, voltado para trabalhadores da informação que atuam nas organizações, traz um conjunto de questões analíticas que não requer um conhecimento prévio em conceitos da ciência da informação ou da computação, podendo ser aplicado a um escopo amplo: tanto de dados (SGBD) quanto de informação (RI).

A justificativa para artefatos como o desenvolvido nesta pesquisa, está no entendimento da sociedade contemporânea, denominada de sociedade da informação, caracterizada pelo processo crescente de digitalização, decorrente da computação impregnada (*pervasive computing*) à sociedade como um todo. Este cenário favorece a ampla oferta de acervos de dados e, a partir destes, a geração de informações. O aumento da oferta de acervos traz consigo uma nova realidade bastante distinta do final do século passado: da falta de dados e informações para a grande oferta e variedade de acervos. Neste sentido, informações como as apresentadas no artefato proposto nesta pesquisa, constituem-se como importante subsídio ao desenvolvimento de métodos e ferramentas de apoio ao processo de tomada de decisão, associado à seleção de acervos. Em um modelo racional para tomada de decisão, como o *Analytic Hierarchy Process* (AHP) (Saaty, 1980), as dimensões de qualidade descritas no artefato podem ser consideradas como critérios, os procedimentos para análise e mensuração destes, como o meio de atribuir um valor a estes critérios. Muitos dos procedimentos propostos para análise ou mensuração da qualidade das dimensões, não retornam números para aplicação direta em uma equação de tomada de decisão como no modelo AHP, ou seja, há a necessidade de esforços e trabalhos de pesquisa complementares neste sentido.

Como já destacado na seção de método, esta pesquisa concentrou-se, predominantemente, nos “testes iniciais do artefato” (fase 4, da figura 1). A utilização do artefato por cinco praticantes de um grupo de pesquisa, não pode ser considerado como uma “validação do artefato no contexto da área fim” (fase 5, da Figura 1). Para que o teste ocorrido fosse considerado como um teste de campo válido, os cinco pesquisadores deveriam ter testado todas as questões do artefato, inclusive as iniciais

pertinentes as de pré-análise dos acervos, associadas às dimensões contextualização e abrangência do acervo em relação ao tema que se deseja obter mais conteúdo. Desta forma, esta é uma limitação da presente pesquisa, como também se configura como uma oportunidade para continuidade da presente pesquisa: realizar testes de campos com o artefato, ou seja, utilização efetiva do artefato proposto por trabalhadores da informação interessados em analisar a qualidade de diferentes acervos.

REFERÊNCIAS

- Batini, C., et al. (2009). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*, 41(3), 16.1-16.52.
- Bawden, B., & Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180-191.
- Bettman, J.R. (1979). *An information processing theory of consumer choice*, Reading, MA: Addison-Wesley.
- Canales, M., & Peinado, A. Grupos de Discusión. (1994). In: Delgado, J.M., Gutiérrez, J. (Org.) *Métodos y Técnicas cualitativas de investigación em Ciencias Sociales*. Madrid: Síntesis, 288-316.
- Cunha, M.B. (1989). Bases de Dados no Brasil: um potencial inexplorado. *Ciência da Informação*, 18(1), 45-57.
- Damanpour, F., Gopalakrishnan, S. (2001). The dynamics of the adoption of product and process innovations in organizations. *Journal of Management Studies*, 38, 45–65.
- Date, C.J. (2004). *Introdução a Sistemas de Banco de Dados*. São Paulo: Elsevier Editora.
- Davenport T. H. (1993). *Process innovation: reengineering work through information technology*. Boston: Ernst & Young.
- De Sordi, J.O. (2013). *Elaboração de Pesquisa Científica: Seleção, Leitura e Redação*. São Paulo: Editora Saraiva.
- Dosi, G. (1988). The nature of the innovative process. In: Dosi, G. et al. (Org.) *Technical change and economic theory*. Londres: Pinter.
- Eppler, J.M., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and related Disciplines. *The Information Society*, 20(5), 325-344.
- Gao, J. et al. (2010). Solving the Information Overload Problem: The Role of Unconscious Thought in Enhancing Online Purchasing Decisions. *Proceedings do Pacific Asia Conference on Information Systems* (PACIS 2010), Taipei, Taiwan, July 2010. Retirado de: <http://aisel.aisnet.org/pacis2010/129>
- Garvin, D.A. (1988). *Managing Quality*. New York: The Free Press.

- Gaver, W.W. (1996). Situating action II: Affordances for interaction - the social is material for design. *Ecological Psychology*, 9, 111–129.
- Gorey, R.M., & Dobat, D.R. (1996). Managing inthe Knowledge Era. *The Systems Thinker*, Waltham, 7(8), 1-5.
- Gregor, S., & Hevner, A.R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337-355.
- Hevner, A.R. et al. (2004). Design science in Information Systems Research.*MIS Quarterly*, 28(1), 75-105.
- Huang, K.Y. et al. (1999). *Quality Information and Knowledge*. Upper Saddle River, NJ: Prentice-Hall.
- Iivari, J. (2007). A paradigmatic analysis of information systems as a design science.*Scandinavian Journal of Information Systems*, 19(2), 39-64.
- International Center for Media and the Public Agenda. (2007). *Openness & Accountability: A Study of Transparency in Global Media Outlets*. Retirado de: <<http://www.icmpa.umd.edu/pages/studies/transparency/main.html>> Acesso em: 27 jul. 2014.
- Kuechler, B., & Vaishnavi, V. (2008). Theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, 17(5), 489–504.
- Kuo, F. et al. (2013). A Study of Social Information Control Affordances and Gender Difference in Facebook Self-Presentation. *Cyberpsychology, Behavior & Social Networking*, 16(9), 635-644.
- Lazarte, L. (2000). Ecologia cognitiva na sociedade da informação. *Ciência da Informação*, 29(2), 43-51.
- Leifer, R. et al. (2003). *Radical innovation: how mature companies can outsmart upstarts*. Boston: Harvard Business School Press.
- Lino, R. (2005). *Metodologia Científica*. 3. ed.São Paulo: Edições Loyola.
- Loshin, D. (2011). *The practitioner's guide to data quality improvement*. Burlington, MA: Morgan Kaufmann.
- Ma, X. J. et al. (2010). Information explosion on complex networks and control. *European Physical Journal B - Condensed Matter*, 76(1), 179-183.
- Niosi, J. et al. (1993). National systems of innovation: in search of a workable concept. *Technology in Society*, 15(2), 207-227.
- Roper, S., & Love, J. H. (2002). Innovation and export performance: evidence from UK and german manufacturing plants. *Research Policy*, 31, 1087-1102.
- Rybczynski, T. (2007). UC For all employees transforms the enterprise. *Business Communications Review*, 37(6), 30-34.
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Salmela, H. (1997). From information systems quality to sustainable business quality. *Information and Software Technology*, 39(12), 819-825.

Shane, S. A. (2003). *A general theory of entrepreneurship: the individual-opportunity nexus*. Massachusetts: Edward Elgar Publishing.

Smith, M. (2002). DSpace: An Open Source Institutional Repository for Digital Material. *D-Lib Magazine*, 8(10).

Sundbo, J., & Gallouj, F. (2000). Innovation as a Loosely Coupled System in Services. In: Metcalfe, J., Miles, I. (Orgs.) *Innovation Systems in the Service Economy: Measurement and Case Study Analysis*. Boston: Kluwer Academic Publishers, 43-68.

Tidd, J. et al. (2008). *Gestão da Inovação*, ed. 3, Porto Alegre: Bookman.

Tozer, G. (1999). *Metadata Management for Information Control and Business Success*. Norwood, MA: ArtechHouse.

Wilder Jr., J. W. (1981). *New concepts in technical trading systems*. New York: Trends Research.

INNOVATION IN PROCESS OF DIGITAL CONTENT ANALYSIS AND SELECTION

ABSTRACT

The pervasive computing is one of the characteristics of the information society, which provides greater diversity and availability in offering collections of data and information. Currently the informational problem is not the lack of data and information, but its excess, which makes complex and lengthy process of analysis of the quality and selection of digital content, is such contents in databases or information repositories. Aiming to improve this process, this paper proposes an artifact that makes the analysis of the quality of digital collections. This artifact is innovative by direct actions on the basis of priorities and considers specific databases and information repositories aspects. The research is qualitative and applied the technique of group discussion, covering the initial stages of the design science research (DSR), phases associated with the design and the affordance of an artifact of a scientific nature. Five researchers were trained in the use of the artifact, which used it to analyze five collections associated with micro and small Brazilian companies, being three data available in the database and two data available at information repositories. The proposed device was evaluated using a scale of opinions and attitudes. The processes of research contributed to the evolution of the artifact, identifying and implementing improvement points that made the most functional artifact according to its users and more suitable for the next stage of DSR, which implies the utility test.

Keywords: Repository of information; Database; Quality of information; Quality of data.

Data do recebimento do artigo: 29/03/2014

Data do aceite de publicação: 20/08/2015

Apêndice A – Acervos utilizados para os testes do artefato

Cinco acervos foram utilizados como insumos para os testes do artefato, tanto dos realizados pelos próprios desenvolvedores (“teste de laboratório”), quanto pelos usuários finais (“teste de campo”). As bases de dados dos três bancos de dados e os relatórios dos dois repositórios de informação, utilizados nos testes, estão descritos no quadro abaixo.

Banco de Dados 1 (BD-1): Instituto Brasileiro de Geografia e Estatística (IBGE)

O IBGE se constitui no principal provedor de dados e informações do País, que atendem às necessidades dos mais diversos segmentos da sociedade civil, bem como de órgãos das esferas governamentais federal, estadual e municipal.

Bases de dados(do banco de dados do IBGE consideradas nas análises):

- Cadastro Central de Empresas
- Tabela 987 - Empresas e outras organizações, pessoal ocupado total e assalariado em 31.12, salários e outras remunerações e salário médio mensal, por seção, divisão e grupo da classificação de atividades (CNAE 2.0) e faixas de pessoal ocupado total.

URL:www.ibge.gov.br

Banco de Dados 2 (BD-2): Ministério do Trabalho e Emprego (MTE)

A gestão governamental do setor do trabalho conta com o importante instrumento de coleta de dados denominado de Relação Anual de Informações Sociais (RAIS). Instituída pelo Decreto nº 76.900, de 23/12/75, a RAIS tem por objetivo: suprir às necessidades de controle da atividade trabalhista no país, prover dados para a elaboração de estatísticas do trabalho, e disponibilizar informações do mercado de trabalho às entidades governamentais.

Bases de dados (do banco de dados do MTE consideradas nas análises):

- Relação Anual de Informações Sociais (RAIS)

URL:www.mte.gov.br

Banco de Dados 3 (BD-3): Fundação Sistema Estadual de Análise de Dados (SEADE)

O SEADE é uma fundação vinculada à Secretaria Estadual de Planejamento e Desenvolvimento Regional do Estado de São Paulo, um centro de referência nacional na produção e disseminação de análises e estatísticas socioeconômicas e demográficas. Para isso ela realiza pesquisas diretas e levantamentos de informações produzidas por outras fontes, compondo um amplo acervo, disponibilizado gratuitamente, que permite a caracterização de diferentes aspectos da realidade socioeconômica e evolução histórica do Estado, de suas regiões e de seus municípios.

Bases de dados (do banco de dados do SEADE consideradas nas análises):

- Pesquisa de Atividade Econômica Paulista (PAEP)

URL:www.seade.gov.br

Repositório de Informações (RI-1): Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE)

Por meio de parcerias com os setores público e privado, o SEBRAE promove programas de capacitação, estímulo ao associativismo, desenvolvimento territorial e acesso a mercados. Trabalha pela redução da carga tributária e da burocracia para facilitar a abertura de mercados e ampliação de acesso ao crédito, à tecnologia e à inovação das micro e pequenas empresas.

Relatórios(do repositório de informação do SEBRAE considerados nas análises):

- Anuário do Trabalho na Micro e Pequena Empresa

URL: www.sebrae.com.br

Repositório de Informações (RI-2): Federação do Comércio de Bens, Serviços e Turismo do Estado de Minas Gerais (FECOMERCIO-MG)

A FECOMERCIO-MG orienta, coordena, protege, defende e representa legalmente as atividades e categorias econômicas do comércio de Minas Gerais. A entidade também presta serviços de assessoria empresarial, jurídica e econômica para empresas do comércio de bens, serviços e turismo.

Relatórios (do repositório de informações do FECOMERCIO-MG considerados nas análises):

- Pesquisa de Sondagem de Opinião
- Balanço de Crédito do Comércio Varejista de Belo Horizonte
- Pesquisa de Opinião do Comércio Varejista de Belo Horizonte
- Pesquisa de Opinião do Comércio Varejista
- Sondagem de Opinião do Lojista

URL: www.fecomerciomg.org.br