

AS DIFERENTES DIMENSÕES DO REUSO DE DADOS CIENTÍFICOS

THE DIFFERENT DIMENSIONS OF SCIENTIFIC DATA REUSE

Renata Gonçalves Curty

renatacurty@uel.br

Universidade Estadual de Londrina

Resumo: Os avanços da eScience e da ciberinfraestrutura que lhe dá sustentação vêm permitindo que dados científicos sejam compartilhados e acessados de forma mais aberta. Políticas mandatórias de agências de fomento e de periódicos científicos para o compartilhamento de dados têm contribuído para o aumento da disponibilidade de dados de pesquisa por meio de repositórios de dados, criando condições favoráveis para a reutilização de dados para fins nem sempre esperados pelos coletores originais. Apesar dos esforços atuais para promover a transparência e reprodutibilidade na ciência, a reutilização de dados não pode ser presumida, nem considerada uma atividade estimulada meramente pela parcimônia. A inexistência de uma visão integrada dos fatores individuais, sociais e tecnológicos que influenciam a intenção e o comportamento de reuso de dados científicos foi a principal motivação para esta pesquisa. Entrevistas realizadas com 13 cientistas permitiram identificar 25 fatores que influenciam suas percepções e experiências, tanto mal, quanto bem-sucedidas no que tange o reuso de dados produzidos por outros pesquisadores. Esses fatores foram agrupados em seis variáveis teóricas: 1) benefícios percebidos; 2) riscos percebidos; 3) esforços percebidos; 4) reusabilidade dos dados; 5) condições facilitadoras, e 6) influência social. Os resultados obtidos oferecem contribuição teórico-prática para a comunicação científica, pois propiciam uma compreensão ao mesmo tempo ampla e aprofundada sobre os fatores que influenciam o reuso de dados de pesquisa no contexto da ciência aberta, bem como provêm indicadores para políticas e iniciativas que visem tornar dados científicos ativos científicos com maior potencial de reuso.

Palavras-chave: Comunicação Científica. Dados Científicos. Reuso de Dados. Ciência Aberta. Ciências Sociais.

Abstract: The advancement of eScience and its cyberinfrastructure have enabled data to be shared and accessed more openly. Funding agencies and journals' mandates for data sharing have contributed to the increasing availability of research data through data repositories, which create favorable conditions for the reuse of data for purposes not always anticipated by original collectors. Despite the current efforts to promote transparency and reproducibility in science, data reuse cannot be assumed, nor merely considered a "thrifting" activity where scientists shop around in data repositories considering only the ease of access to data. The lack of an integrated view of individual, social and technological influential factors to intention and actual data reuse behavior was the key motivator for this study. Interviews with 13 scientists produced 25 factors that were found to influence their perceptions and experiences, including both their unsuccessful and successful attempts to reuse data. These factors were grouped into six theoretical variables: 1) perceived benefits; 2) perceived risks; 3) perceived effort; 4) perceived reusability; 5) facilitating conditions and, 6) social influence. Research findings offer an important contribution to scholarly communication studies by providing an in-depth understanding about the reuse of research data in the context of open science which can be valuable in terms of theory and practice to help leveraging data reuse and making publicly available data more actionable.

Keywords: Scholarly Communication. Scientific Data. Data Reuse. Open Science. Social Sciences.

INTRODUÇÃO

Dados primários têm sido considerados a moeda mais valiosa da ciência (DAVIS; VICKERY, 2007). A ampla disponibilidade e acessibilidade dos dados da investigação são itens fundamentais da agenda da ciência aberta, que visa maximizar a relação de custo-eficácia dos recursos socioeconômicos, aumentar a utilidade e aplicação de dados para além do foco ou das limitações de tempo dos coletores de dados originais, e promover um melhor controle, reprodutibilidade e transparência na ciência (FIENBERG; MARTIN; STRAF, 1985).

Em busca da ampla disponibilidade de dados de pesquisa e do cumprimento das novas diretivas governamentais, uma série de agências de financiamento, editores de periódicos científicos, instituições acadêmicas e organizações de pesquisa estadunidenses iniciaram uma chamada intensiva para o compartilhamento de dados primários de pesquisa nos últimos anos (XIA et. al., 2012).

À medida que novos repositórios de dados são criados para abrigar os dados da pesquisa, e mais dados se acumulam em seus servidores, a atenção se desloca para encontrar maneiras de sustentar o valor destes resultados da investigação e maximizar a sua reutilização (FANIEL; ZIMMERMAN, 2011). Isto porque os benefícios do compartilhamento de dados só podem ser mensurados por meio de seu reuso efetivo (NIU, 2009b). Neste sentido, a sustentabilidade do ciclo de vida da ciência aberta depende de condições propícias para a maximização do reuso de dados científicos, ao invés de meramente acumulá-los ociosamente em repositórios de dados.

Tempo, dinheiro e economia de esforço são frequentemente reconhecidos como principais motivadores para o reuso de dados de pesquisa (CASTELO, 2003; HYMAN, 1972; KIECOLT; NATHAN, 1985; LAW, 2005). Embora a parcimônia seja fator preponderante, este estudo argumenta que o reuso de dados não pode ser reduzido à obtenção de “dados de segunda-mão”, e mais “baratos”. O reuso de dados científicos é um processo mais complexo que exige que os cientistas tenham capacidade de descobrir e identificar dados confiáveis e relevantes (THESSSEN; PATTERSON, 2011). Além disso, e talvez mais importante, o reuso exige que cientistas sejam capazes de traduzir e re-contextualizar os dados primários obtidos por

outros pesquisadores, a fim de aplicar para os seus próprios fins, sem que haja má-interpretação ou má-utilização destes.

Ao reutilizar dados de pesquisa disponíveis publicamente, cientistas enfrentam a dualidade entre a conveniência de ter acesso a dados prontos e o esforço de lidar com dados produzidos por outra pessoa. Se por um lado trabalhar com dados existentes tem como vantagem a redução de custos e tempo associados à coleta de dados (CASTLE, 2003; LAW, 2005), por outro lado o reuso impõe limitações aos pesquisadores que terão de lidar com dados criados em circunstâncias particulares, por meio da adoção de procedimentos e técnicas específicas de coleta de dados, a fim de responder a questões pontuais de investigação (BOSLAUGH, 2007; DEVINE, 2003). Tais dificuldades acontecem pelo fato de dados científicos não serem commodities flutuantes e desvinculadas dos contextos e relações nos quais foram produzidos (MAUTHNER, 2012).

A despeito do reconhecimento desta dualidade, pouco se sabe sobre como os cientistas percebem o processo de reuso de dados e os diferentes fatores que os motivam e/ou desencorajam a reutilizar dados coletados por outros pesquisadores (WALLIS; ROLANDO; BORGMAN, 2013; ZIMMERMAN, 2007; 2008). Ainda menos é sabido sobre como esses fatores afetam as intenções e o reuso efetivo de dados científicos. Apesar do amplo reconhecimento da importância do reuso de dados na ciência, este problema tem sido abordado pela literatura de forma ainda periférica, como um resultado desejável e esperado das iniciativas e políticas para o compartilhamento de dados. A falta de uma visão mais integrada quanto aos diferentes fatores que afetam coletivamente o reuso de dados motivou o estudo relatado neste artigo, que tem como objetivo oferecer uma visão mais clara dos diferentes fatores (individuais, sociais e tecnológicos) capazes de influenciar o reuso de dados publicizados via repositórios digitais.

2 REUSO DE DADOS CIENTÍFICOS: CONCEITOS E ABORDAGENS

O reuso de dados científicos pode ser definido como toda nova aplicação de dados por meio de re-análise e replicação, ou como a combinação de diferentes conjuntos de dados por meio de integração ou meta-análise, a partir de novas perguntas de pesquisa, novos métodos

de análise, com propósitos similares ou distintos daqueles empreendidos no estudo original, com ou sem a participação do reutilizador (CURTY, 2015).

Existem apenas alguns estudos empíricos na literatura acerca do comportamento e/ou da percepção de cientistas com relação ao reuso de dados coletados/produzidos por outrem. Estes estudos podem ser agrupados em duas categorias com base na abordagem de pesquisa adotada. Uma primeira examina o comportamento reutilização de dados por meio de pesquisa documental e da análise de citações de coleções de dados (*datasets*), enquanto a segunda investiga empiricamente o comportamento de reuso e/ou a percepção dos cientistas quanto o reuso.

O primeiro grupo de estudos tenta explicar os dados dos cientistas comportamento de reuso por análise bibliométrica, incluindo os estudos de Piwowar (2008; 2010ab), Piwowar e Vision (2013) e Chao (2011, 2012), que rastrearam citações de conjuntos de dados em publicações nas áreas Biomédicas e das Ciências da Terra. Em suma, estes estudos consideram que as citações e a atribuição de autoria para os conjuntos de dados são medidas fiáveis que, ao serem rastreadas, permitem demonstrar certos padrões de comportamento de reuso de dados científicos. Por exemplo, Piwowar e Vision (2013) descobriram que os cientistas que reutilizam dados abertos tendem a ser mais citados do que aqueles que publicam artigos baseados em seus próprios conjuntos de dados.

A análise de citação de *datasets* tem permitido mapear quais coleções de dados têm sido utilizadas, as redes de reuso, bem como mensurar comportamentos de reuso entre cientistas individuais e de grupos, tanto em disciplinas específicas, quanto em circunstâncias inter e transdisciplinares. A abordagem de análise de citação, no entanto, tem como limitação a impossibilidade de explicar e de capturar as diferentes nuances do processo de decisão pelo reuso e do comportamento de reuso de dados propriamente.

O segundo grupo de estudos adota abordagem majoritariamente qualitativa para investigar as percepções, experiências e atitudes dos cientistas quanto ao reuso de dados (CARLSON; ANDERSON, 2007; FANIEL; JACOBSEN, 2010; HOWARD et al., 2010; ZIMMERMAN, 2003; 2007). Porém, tais estudos exploratórios sobre o comportamento de reuso de dados são fundamentalmente atóricos e/ou conduziram a análise em um nível ainda elementar de

abstração, o que torna difícil identificar variáveis ou hipóteses de pesquisa que possam fundamentar ou direcionar pesquisas futuras acerca do reuso de dados científicos.

Devido à novidade do tema e à incipiência da literatura acerca deste fenômeno, este artigo reporta um estudo que buscou elucidar e compreender as diferentes dimensões do reuso de dados científicos. Ao explorar as experiências dos cientistas sobre o reuso de dados de pesquisa, este estudo teve como objetivo analisar como os cientistas avaliam a possibilidade de reuso de dados coletados por outros investigadores e quais os fatores eles julgam determinantes para o reuso de dados, com vistas à proposição de um modelo conceitual preliminar que possa nortear pesquisas futuras nas searas da sociologia da ciência e da comunicação científica.

3 PROCEDIMENTOS METODOLÓGICOS

A população-alvo do estudo foi constituída por cientistas sociais estadunidenses. A opção pelas Ciências Sociais foi motivada pela possibilidade obter mais nuances e variações no processo de reuso de dados em comparação a outras disciplinas. As Ciências Sociais incluem uma vasta gama de disciplinas, e essa diversidade é entendida como particularmente útil para fornecer exemplos de práticas de reuso de dados heterogêneos, obtidos por métodos científicos diversos, fundados em diferentes tradições disciplinares. Além disso, os dados nas Ciências Sociais são tidos como intensivamente contextuais e dependentes do tempo de coleta, o que requer esforço extra dos que os reutilizam para preservar a interconexão de dados e reflexividade necessária, de modo a garantir a compreensibilidade e o valor informativo dos dados reutilizados (FRIEDHOFF et al., 2013; JACOBY, 2010).

O processo de recrutamento de participantes para o estudo teve como requisito principal a familiarização com o processo de reuso de dados. Neste sentido, os sujeitos potenciais deveriam ter ao menos empreendido tentativa de reuso de dados primários gerados por outros pesquisadores, em suas próprias pesquisas; independente do resultado final ter resultado no reuso efetivo.

A amostragem intencional não-probabilística foi aplicada a fim de recrutar potenciais participantes com características relevantes para a pesquisa. A fase de recrutamento foi

consentida, apoiada e mediada pela equipe dos principais repositórios de dados nas Ciências Sociais hospedados nos Estados Unidos: o *Interuniversity Consortium for Political and Social Research* (ICPSR) e o *Harvard Dataverse Network*, equipes estas que enviaram e-mails solicitando participação voluntária na pesquisa a todos os usuários cadastrados. Após três rodadas de solicitação de participação, 13 cientistas sociais (sete homens e seis mulheres) foram entrevistados.

Na ocasião da pesquisa os participantes eram filiados a diferentes instituições acadêmicas e de pesquisa localizadas nos estados de Nova York, Oregon, Massachusetts, Pensilvânia e Mississippi; e pertenciam a diferentes disciplinas, incluindo economia, ciências políticas, sociologia, comunicação e mídia social, estudos da criança e da família, psicologia clínica, administração pública, e assuntos internacionais.

A pesquisa de caráter exploratório e abordagem qualitativa adotou um protocolo de entrevista semiestruturado com perguntas elaboradas que seguiram a estrutura de funil. Para uma melhor compreensão e interpretação das narrativas dos participantes, os participantes foram inicialmente questionados sobre questões mais gerais, como suas agendas de pesquisa e áreas de estudo, bem como com relação às suas perspectivas quanto o reuso de dos dados na ciência de modo geral, e na disciplina a que pertencem. Em seguida, a entrevista adotou questões mais específicas sobre as experiências de reuso dos participantes.

Todas as entrevistas foram gravadas para a conveniência de transcrição, totalizando sete horas e 55 minutos de gravação de áudio. Dada à dispersão geográfica dos participantes e suas preferências de comunicação, cinco entrevistas foram realizadas face-a-face, cinco por meio de chamadas por vídeo via *Skype*, e três por telefone.

As entrevistas foram transcritas utilizando o software *Express Scribe* e as transcrições foram posteriormente registradas no software QSR NVivo 10 para a conveniência da organização, codificação e análise dos dados. O processo de análise seguiu as cinco etapas da análise de conteúdo, a saber: 1) preparação das informações; 2) unitarização do conteúdo e das unidades de análise; 3) categorização das unidades; 4) descrição, e 5) interpretação (MORAES, 1999).

As entrevistas foram codificadas seguindo um processo de análise indutiva (*bottom-up*) sem seguir, portanto, qualquer esquema preconcebido de codificação. No entanto, a análise esteve atenta aos aspectos de exaustividade, homogeneidade, exclusividade e objetividade necessários para a fidedignidade da interpretação dos dados.

Inicialmente a codificação teve como foco alguns padrões nas respostas quanto às motivações de desmotivações para o reuso de dados. Diferentes rodadas de análise minuciosa das transcrições foram realizadas para identificar temas emergentes e ocorrências semelhantes, tanto nas entrevistas individuais, quanto no conjunto de entrevistas. A análise de dados permitiu identificar 25 fatores importantes para a compreensão das motivações e desmotivações que influenciam a intenção de reuso e o comportamento de reuso de dados nas Ciências Sociais, conforme será detalhado na seção seguinte.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

As entrevistas revelaram um conjunto de fatores que influenciam as práticas de reuso de dados entre cientistas sociais. Seis dimensões principais (variáveis teóricas) foram criadas para representar os 25 diferentes fatores que totalizaram 430 repetições nas transcrições, sendo elas: 1) benefícios percebidos, 2) riscos percebidos, 3) esforços percebidos, 4) influência social, 5) condições facilitadoras, e 6) reusabilidade dos dados. Conforme síntese apresentada no APÊNDICE A, a coluna “N” indica o número de entrevistas de ocorrência dos fatores, e a coluna “R” informa o total de repetições destes fatores nas diferentes entrevistas. O APÊNDICE A também apresenta definição sucinta dos fatores para fins da pesquisa, seguida de literatura relacionada que serviu de referência para a triangulação e análise dos dados. A identidade dos sujeitos de pesquisa foi preservada com a adoção de pseudônimos, e os trechos selecionados para ilustrar argumentos da análise foram traduzidos do original em inglês para a língua portuguesa.

4.1 BENEFÍCIOS PERCEBIDOS

Os benefícios percebidos representam fatores mencionados pelos entrevistados como motivações pessoais ou que os cientistas sociais acreditam que sejam relevantes para a

comunidade científica em geral, quando optam pelo reuso de dados existentes, ao invés de empreenderem nova coleta de dados.

O primeiro fator nesta categoria, a “Expansão do Conhecimento” corresponde à possibilidade de cientistas produzirem novas descobertas científicas e contribuírem para o desenvolvimento de um campo particular por meio do reuso de dados. Este aspecto converge com a ideia de Hyman (1972) que postula que o reuso de dados expande os horizontes intelectuais dos investigadores e, conseqüentemente, o seu campo de estudo. Assim, os pesquisadores são estimulados a pensar sobre os problemas de pesquisa sob novas perspectivas e em um maior nível de abstração (HYMAN, 1972). Uma ideia semelhante foi expressa por Frank ao afirmar que:

(...) Alguém pode olhar para os dados em outro momento, em algum outro lugar, talvez com completamente diferentes propósitos e objetivos ou ferramentas que outra pessoa, e obter novas informações nestes dados antigos. Mas ao analisá-los de uma maneira diferente, temos novas informações (...), você sabe, algo aplicável, uma nova ideia de como usar esses dados antigos (Frank).

Um segundo fator relacionado aos benefícios percebidos é o da “parcimônia”. Cientistas sociais percebem a reutilização reuso de dados como uma forma de contornar os problemas associados com a coleta de dados primários, incluindo a redução do tempo e de esforço necessário para a obtenção de dados. Os cientistas sociais veem o reuso de dados como benéfico, pois se trata de uma oportunidade de obtenção de dados que nem sempre são de fácil coleta.

A noção de frugalidade como um benefício e impulsionador associado ao reuso de dados tem sido extensivamente abordada na literatura. Kiecolt e Nathan (1985) articulam sobre a análise de dados secundários como uma atividade de economia de recursos. Hyman (1972) enfatiza que o reuso de dados científicos existentes reflete em economiza dinheiro, tempo e pessoal em pesquisas. De modo semelhante, Law (2005) e Castle (2003) elaboram sobre a questão da parcimônia e sobre a possibilidade de pesquisadores iniciantes ou com poucos recursos contarem com dados coletados por instituições e pesquisadores mais experientes, conforme pode ser evidenciado no comentário a seguir:

É preciso um grande esforço para coletar dados, e como um pesquisador experimental eu sei que, certamente, a disponibilidade de dados já existentes pode fazer todo o processo de pesquisa mais ágil (Cindy)

Por fim, um terceiro fator associado aos benefícios percebidos é a noção de “pré-endossamento”. Cientistas sociais percebem a reutilização dos dados como benéfica, pois os dados disponíveis para reutilização são considerados, de alguma forma, credíveis e fiáveis, caso contrário, não seriam compartilhados e disponibilizados de forma aberta e transparente, para estarem sujeitos ao escrutínio e verificação. Esta percepção de endossamento prévio pode ser verificada no depoimento de Ellen:

A maioria dos dados que estão no ICPSR ou em outros repositórios (...) provavelmente tem subsídios, foram avaliados por seus pares, e eles são julgados, até certo ponto (Ellen).

4.2 RISCOS PERCEBIDOS

Os riscos percebidos são considerados consequências potencialmente prejudiciais associadas ao reuso de dados de pesquisa. Quatro tipos de riscos foram encontrados nas narrativas dos entrevistados.

A primeira, “desvalorização”, diz respeito à possibilidade dos cientistas que reutilizam dados receberem menos crédito em comparação aos cientistas que realizaram a coleta dos dados primários (originais) em suas pesquisas. Goodwin (2012) elabora sobre este assunto, indicando que os cientistas sociais, especialmente com uma abordagem de inquérito qualitativa, tendem a subestimar o reuso de dados. Martin (1995) e Fahs, Morgan e Kalman (2003) indicam que, especialmente em casos de estudos de replicação, há uma sensação de que o reuso dos dados pode não ser considerado como uma atividade de pesquisa. Esta questão ficou patente na narrativa de Denise:

[...] havia um problema de valor [atribuído à minha pesquisa] acontecendo (...). Existia realmente uma questão de valor e [minha pesquisa] não foi tão respeitada como se eu tivesse coletado meus próprios dados (Denise).

Uma segunda categoria associada aos riscos percebidos é a possibilidade de “infração de códigos éticos”. Os cientistas sociais podem hesitar em reutilizar dados gerados por outros

pesquisadores com receio, de inadvertidamente, violarem questões éticas em pesquisa. Quando a pesquisa envolve dados sensíveis, o consentimento informado não pode ser simplesmente presumido pelos reutilizadores e há uma necessidade de verificação se o reuso de dados viola o contrato estabelecido entre os sujeitos e os investigadores principais. Além disso, questões de direitos autorais e de confidencialidade podem não ser claras para os reutilizadores potenciais (HEATON, 2004).

Grinyer (2009) e Law (2005) indicam que este risco pode ser resultado da falta de clareza com relação aos códigos de conduta ética, especialmente para os dados qualitativos arquivados em repositórios para reuso futuro; preocupação também expressa por Ellen:

Existem alguns conjuntos de dados que exigem, você sabe (...) confidencialidade, eu diria, que têm requisitos de confidencialidade e você precisa descobrir se esses conjuntos de dados requerem mecanismos de confidencialidade para que possa reutilizá-los (Ellen).

O terceiro fator de risco diz respeito ao potencial “deslize”. As entrevistas revelaram que cientistas sociais se preocupam com a possibilidade, do reuso de dados incorrer na má interpretação, ou na utilização incorreta e inadvertida dos dados originais. Kuula (2010) descreve o uso indevido dos dados como uma das principais preocupações dos cientistas sociais. Corti e Thompson (2004) explicam que "preocupações sobre a má interpretação dos dados podem surgir do receio quanto à interpretação seletiva e oportunista na reanálise" (p. 307, tradução nossa). Em outras palavras, o processo de análise dos dados gerados por outros sob diferentes ângulos pode gerar suposições erradas, pois os dados são extraídos do seu quadro inicial contextual. Tenopir et al. (2011) identificaram a má interpretação e uso indevido de dados como preocupações recorrentes dos cientistas ao considerarem o compartilhamento e o reuso de dados em suas pesquisas, preocupações estas também ressaltadas por Cindy:

Eu tenho as mesmas preocupações com relação ao reuso dos meus dados, assim como os que eu reutilizo. Eu certamente quero saber se eles [os pesquisadores que reutilizam meus dados] estão seguindo certos padrões e coisas assim, e que eles não estão fazendo mau uso deles. (Cindy).

O quarto e último fator relacionado aos riscos percebidos corresponde à possibilidade da “vulnerabilidade a erros”. Dados coletados, compilados e organizados por outros podem conter erros ocultos que não são facilmente identificáveis pelos reutilizadores. A ideia de erros escondidos é articulada por Kiecolt e Nathan (1985) como um risco potencial para os reutilizadores. Da mesma forma, Hyman (1972) afirma que durante a realização de análise de dados de investigação existentes, cientistas que reutilizam dados enfrentam dificuldades na detecção de erros, pois há uma falta de controle dos reutilizadores sobre a qualidade dos dados (CASTLE, 2003).

4.3 ESFORÇOS PERCEBIDOS

Os esforços percebidos representam a energia que os cientistas sociais estimam investir para lidar com dados produzidos/coletados por outros pesquisadores. A análise das entrevistas permitiu identificar seis diferentes fatores relacionados à dimensão dos esforços percebidos.

O primeiro fator de esforço “inovar com dados antigos” diz respeito à necessidade dos cientistas de enveredar esforços na identificação de novas formas de abordar dados antigos, diferentes daquelas empreendidas na análise da pesquisa original, e dos reusos subsequentes que os dados tenham possivelmente sofrido. Zimmerman (2008) discute a ideia de novos conhecimentos a partir de dados antigos. A autora descobriu que cientistas não apenas devem se atentar para os dados que têm em mãos, mas também procurar formas de expandir a ciência e as possibilidades de análise dos dados a serem reutilizados. Esta noção de esforço associado com a necessidade de buscar o novo em dados antigos fica evidente no comentário de Michael:

Quando se trata de conjuntos de dados utilizados publicamente às vezes tenho a sensação de que [o reuso] é um trabalho intenso, porque [outros pesquisadores] já extraíram diferentes perguntas sobre aqueles dados (...) por isso há, pelo menos, essa percepção de que estou competindo com um monte de pessoas pelos mesmos dados. (Michael).

O segundo fator nesta dimensão relaciona-se ao processo de obtenção de “acesso aos dados”. O acesso ao conjunto de dados varia dependendo de quem possui o controle de acesso, onde os dados são mantidos, e em qual formato (HEATON, 2004). O processo de

obtenção de acesso aos dados é reconhecido por cientistas sociais como um dispêndio de esforço na reutilização de dados. Faniel et al. (2013) constataram que a facilidade de acesso aos dados é a variável de maior impacto no reuso de dados. Tenopir et al. (2011) também descobriu que cientistas estão mais propensos e manifestam forte interesse em utilizar conjuntos de dados de outros pesquisadores, quando estes são de fácil acesso.

Outro fator relacionado ao esforço corresponde à descoberta de dados disponíveis para possível reuso. Isto porque os entrevistados enfatizaram a atividade de identificação de dados para potencial reuso uma atividade nem sempre trivial e intuitiva, conforme descrito por um dos entrevistados:

[Localizar dados para reuso] requer fazer chamadas telefônicas, pesquisar intensivamente na web e descobrir onde posso encontrar os dados que preciso para responder à uma pergunta de pesquisa em particular. (Adam).

Faniel e Majchrzak (2002) descobriram que o esforço associado à encontrabilidade de dados é um importante preditor para o reuso de dados. Nesta mesma perspectiva, Darby et al. (2012) enfatizam que a facilidade de descoberta atua diretamente sobre a disposição dos cientistas para reutilizarem dados.

Um quarto fator identificado na dimensão dos esforços percebidos corresponde à necessidade de “lidar com incompatibilidades”. Segundo os entrevistados, a reutilização de dados implica em esforços para lidar com os desequilíbrios entre os dados disponíveis para reuso e os dados que os reutilizadores desejariam ter, a fim de responder às suas perguntas de pesquisa. Isto considerando que dados primários foram coletados em circunstâncias especiais, em um dado contexto e tempo, a fim de investigar questões específicas, o que significa que os dados considerados para reuso raramente captam todos os elementos que os reutilizadores esperam. Kiecolt e Nathan (1985) indicam que pesquisadores, ao reutilizarem dados de pesquisa geralmente experimentam uma incompatibilidade entre os objetivos da pesquisa primária e nova pesquisa. Dale (2004) enfatiza a necessidade de adaptação aos dados disponíveis, pois os analistas secundários (reutilizadores) não têm oportunidade de influenciar as perguntas ou os quadros de codificação aplicados, e este deve ser levado em conta em todas as fases de análise. Quanto a este esforço de ajustes Ivan comentou:

Uma das limitações para aqueles que trabalham com dados secundários é que (...), você não elaborou as perguntas, você não as definiu [...] então podem ser semelhantes ao que você gostaria de saber, mas não exatamente o que você quer saber. (Ivan).

A preparação dos dados primários gerados por outros investigadores foi um quinto fator associado ao esforço percebido identificado e que pode afetar a decisão de cientistas quanto ao reuso. Esta preparação inclui a formatação e recodificação dos dados originais, o gerenciamento de dados inexistentes (em pesquisas quantitativas), bem como o complemento e anotação combinada quando os dados derivam de diferentes fontes e métodos (observações, entrevistas, etc.). Faniel et al. (2013) destacam que a facilidade de operação dos dados desempenha um papel preponderante para a opção pelo reuso de dados existentes. Neste sentido, o reconhecimento de esforços adicionais associados a esta preparação pode inibir o reuso, conforme situação expressa por uma cientista política:

Eu teria que fazer recodificação manual e individual de cada uma das observações na coleção de dados sobre desastres naturais (...). Mesmo como com atalhos, era simplesmente muito trabalho. (Jen).

Por fim, a “compreensão integral do estudo original”, isto é; o estudo a partir do qual os dados primários deriva, constitui a sexta dimensão de esforço identificada na pesquisa. Faniel e Jacobsen (2010) descobriram que cientistas investem uma quantidade significativa de energia buscando a confiança de que os dados sejam completamente compreendidos. Boh (2008) constatou que a complexidade dos dados é um fator importante para a própria intenção de reuso. A autora enfatiza o tempo necessário para que os reutilizadores entendam suficientemente estes dados, a fim de determinar como as ideias podem ser adaptadas para atender o problema de pesquisa que buscam responder. Este fator pode ser constatado na narrativa de Michael:

Lembro-me que era um esforço em si mesmo interpretar o que as perguntas tinham realmente perguntado e em quais momentos. Era necessário fazer um pente fino nos diferentes arquivos da pesquisa original (Michael).

4.4 REUSABILIDADE DOS DADOS

Os entrevistados relataram diferentes atributos que julgam relevantes para determinar o potencial de reuso dos dados. Em outras palavras, os cientistas avaliam que os dados devem de possuir certas características de reusabilidade, sendo neste estudo representadas por cinco fatores distintos, observados nos relatos dos entrevistados.

Um primeiro fator diz respeito à “documentação dos dados”. Cientistas tendem julgar a qualidade da documentação, isto é; o quão o suficiente e completa a documentação acerca dos dados é. A documentação de dados inclui uma variedade de materiais suplementares que têm por função apoiar a compreensão e o reuso dos dados, dentre os quais: dicionários de dados (*codebooks*), relatórios detalhados sobre o processo de coleta de dados, instrumentos de coleta de coletas, publicações anteriores que utilizaram os dados de pesquisa, guias de usuário, documentação de software para extração e análise de dados, e documentos expedidos pelo comitê de ética avaliador. David (1991) enfatiza o papel da documentação no reuso de dados científicos e os riscos causados pela documentação insuficiente. Muitos estudiosos têm abordado a importância da qualidade da documentação como uma condição para o reuso (FANIEL et al, 2013; ZIMMERMAN, 2003; 2007; 2008; NIU, 2009ab; NIU; HEDSTROM, 2009).

Um outro fator identificado foi a “adaptabilidade dos dados” que inclui, dentro de outros aspectos, a adequação ao tema, ao nível de análise, e quão próximos os dados originais são do propósito de reuso (HINDS; VOGEL; CLARKE-STEFFEN, 1997). Palmer, Weber e Crager (2011) consideram a reusabilidade o atributo que determina a utilidade dos dados. Faniel et al. (2013) utilizam o atributo de relevância de dados para representar a aptidão para o reuso.

A “credibilidade dos produtores” de dados também se manifestou nas entrevistas como fator determinante para a avaliação da reusabilidade dos dados. Ao julgar o potencial de reuso de dados científicos, pesquisadores consideram a confiabilidade nas instituições, pesquisadores individuais e/ou grupos de pesquisadores que produziram/coletaram os dados.

Há um extenso corpo de literatura disponível sobre a credibilidade informações. No contexto de dados científicos, Carlson e Anderson (2007) consideram a credibilidade de proveniência dos dados como o fator-chave para a reusabilidade. Darby et al. (2012) ressaltam

a importância de que reutilizadores de dados "sintam-se seguros" sobre os dados de "segunda mão". Com relação à confiabilidade e credibilidade, alguns entrevistados não só destacaram a importância de confiar nos produtores de dados, mas também expressaram suas preferências por conjuntos de dados produzidos por grupos institucionais, ou por pesquisadores individuais filiados a instituições de renome, como ilustrado abaixo:

[...] é mais fácil convencer avaliadores quando os dados provêm de instituições de pesquisa grandes [...]. Quanto à pesquisa individual, você sabe, há uma questão de confiança e creio não ser possível saber ao certo se processos [para obtenção de dados] seguidos pelo pesquisador foram criteriosos. [...]. Eu geralmente prefiro [reutilizar] dados de uma instituição estabelecida (Michael).

A "qualidade dos dados" quanto à completude e à consistência também é avaliada pelos entrevistados quando estes julgam o potencial de reuso dos dados. Faniel et al. (2013) relatam que, após a acessibilidade dos dados, a qualidade de dados é o contribuinte mais forte para a satisfação quanto ao reuso de dados entre cientistas sociais. A qualidade da documentação por vezes se sobrepõe ao julgamento da qualidade dos dados e as distinções entre eles nem sempre são fáceis de capturar (NIU, 2009ab). No entanto, o fator de qualidade de dados reflete as percepções dos entrevistados ao descreverem atributos diretamente ligados ao conjunto de dados (*dataset*), quanto à integralidade e precisão dos mesmos (HINDS; VOGEL; CLARKE-STEFFEN, 1997). A integralidade (completude) dos dados como fator de qualidade determinante para a reusabilidade é expressa na narrativa de Beth:

O primeiro requisito que observo é a completude dos dados. Se eles têm muitos pontos em falta significa que não é um bom conjunto de dados porque há apenas algumas observações (Beth).

Outro fator atrelado à reusabilidade é o "rigor do estudo original". Como forma de ajuizar sobre o potencial de reuso dos dados, os entrevistados informaram que consideram o design metodológico do estudo original, bem como a descrição dos procedimentos do estudo e de sua execução propriamente dita, tal qual recomendado por Hinds, Vogel e Clarke-Steffen (1997). O primeiro indicador aparente de rigor é verificado no próprio conjunto de dados e na documentação de dados que o acompanha. Porém, esta dimensão de reusabilidade diferencia-

se das supracitadas, pois considera o julgamento geral de adequação dos métodos e procedimentos do estudo original.

4.5 CONDIÇÕES FACILITADORAS

As condições facilitadoras englobam os fatores que correspondem à infraestrutura que dá suporte ao reuso de dados, sendo esta variável teórica constituída por cinco fatores observados nas entrevistas.

O primeiro fator diz respeito à “disponibilidade da documentação suplementar”, a qual é destacada por diversos autores (DAVID, 2001; PIGOTT; HOBBS; GAMMACK, 2001; MARKUS, 2001; NIU, 2009ab; NIU; HEDSTROM, 2009; ZIMMERMAN, 2003; 2007; 2008) como condição *sine qua non* para o reuso. Markus (2001) discorre sobre a importância e os desafios de se documentar para o reuso de outras pessoas. Em sua opinião, a documentação para futura reutilização é uma tarefa desafiadora, pois as diferentes aplicações possíveis não podem ser sempre previstas. No entanto, sem a devida documentação destes ativos, as possibilidades de reuso são seriamente comprometidas.

Devido à mais ampla disponibilidade dos dados, alguns entrevistados enfatizaram a importância de ter acesso ou meios para identificar literaturas relacionadas que utilizaram um conjunto de dados em particular. Não só eles acreditam que é útil ter conhecimento de todas as pesquisas que derivam de uma mesma coleção de dados (*dataset*), mas também eles reconhecem que o acesso a estudos relacionados contribui para a identificação de lacunas e/ou potenciais oportunidades de reuso em novas pesquisas. Os achados de Faniel et al. (2013) e de Niu (2009ab) coadunam com os resultados desta pesquisa, pois constata-se que pesquisadores tendem a consultar a literatura relacionada de autoria dos produtores de dados, bem como as produções resultantes de reuso de terceiros, como forma de observar diferentes críticas e abordagens atribuídas ao conjunto de dados de interesse, conforme expresso por Denise:

Eu li muito, li diferentes teses que foram feitas com o mesmo conjunto de dados apenas para ver o tipo de escopo (...) como as pessoas utilizaram o conjunto de dados, de modo que isto foi realmente útil para mim (Denise).

Um segundo fator atrelado às condições facilitadoras corresponde à “disponibilidade de repositórios”. Mecanismos informais e *ad-hoc* para o compartilhamento de dados sempre existiram, mas não são suficientemente eficazes e transparentes considerando o atual contexto da comunicação científica e da ciência aberta. A importância dos repositórios de dados, enquanto infraestrutura tecnológica que centraliza o compartilhamento de dados, preserva, e permite que estes ativos sejam reutilizados, é notória entre acadêmicos (MARKUS, 2001; BORGMAN, 2007; MARCIAL; HEMMINGER, 2010; TENOPIR et al., 2011).

Nas narrativas dos entrevistados, ficou patente que a existência e o conhecimento prévio de repositórios, principalmente que contemplem dados dedicados à disciplina, à temática da pesquisa ou à abordagem metodológica de preferência do pesquisador (qualitativa, quantitativa ou mista), são determinantes para que estes optem pelo reuso de dados. Cindy reforçou o papel dos repositórios de dados ao justificar suas preferências por ainda investir na coleta de dados em suas pesquisas, tendo em vista a inexistência de repositórios dedicados à sua área de investigação:

[Seria importante a existência de repositórios na comunicação] que dados pudessem ser compartilhados, de alguma forma, mesmo a partir de experimentos, você sabe, mais pessoas poderiam fazer certas perguntas e eu poderia olhar para outros itens lá, que seriam realmente valiosos [para as minhas pesquisas] (Cindy).

Outro fator tido como uma condição facilitadora para o reuso de dados consiste na “possibilidade de alcance ou contato com os investigadores primários”. Além da existência dos repositórios, os entrevistados destacaram a importância de terem acesso aos investigadores primários (coletores de dados), a fim de reunir um melhor entendimento das nuances da pesquisa original, que não são facilmente acessíveis e capturadas a partir da análise dos dados e da documentação fornecida. Nesse sentido, Boh (2008) destaca que interações pessoa-a-pessoa entre autores/coletores de facilitam o reuso de dados, especialmente em circunstâncias de dados de alta complexidade. Da mesma forma, Faniel e Zimmerman (2011) reconhecem o valor de troca social entre os produtores de dados e seus reutilizadores, mas indicam que este intercâmbio social é difícil de materializar em grande escala. Apesar disso, alguns entrevistados descreveram o processo de contato com os investigadores principais para solicitar informações

ou esclarecimentos adicionais acerca dos dados que pretendiam utilizar, conforme exemplo a seguir:

Havia alguns detalhes institucionais que eu queria saber [...] e não fomos capazes de encontrar nos arquivos de dados, então eu perguntei a eles sobre alguns dos arquivos que não foram incluídos [no repositório] [...] e eles me forneceram muitos detalhes institucionais que eu precisava (Adam).

O terceiro fator nesta categoria é representado pela “disponibilidade de suporte e assistência”. Markus (2001) e Behboudi e Hart (2008) discutem o papel dos intermediários humanos para o reuso efetivo de dados. Faniel et al. (2013) argumentam que o reuso de dados científicos dependem muitas vezes de assistência humana durante o processo, seja por meio de orientações técnicas, metodológicas e/ou sobre aspectos quanto à proteção dos dados. Produtores de dados podem ser considerados intermediários humanos, mas este terceiro fator abarca o apoio formal provido durante o processo de reuso, o qual pode ser prestado tanto pela instituição na qual o pesquisador que busca reutilizar dados está vinculado, quanto pela equipe do repositório em que o pesquisador obteve acesso aos dados. Nos relatos dos entrevistados fica evidente a importância dos intermediários, representados pela equipe de apoio do repositório, ou por setores como os centros de estatística, bibliotecas, arquivos, e escritórios de ética em pesquisa de suas instituições durante o processo de reuso:

Eu precisava de apoio externo para o processo de análise de preparação dos dados [que gostaria de reutilizar]. Então, você sabe, eu recorri ao centro de estatística do campus e tive ajuda de uma equipe estatística. (Nathan).

O quinto e último fator relacionado à categoria de condições facilitadoras diz respeito à oferta de “treinamentos e programas de formação” que instruem os pesquisadores sobre as possibilidades de reuso de dados e desenvolvam habilidades específicas em suas diferentes manifestações (re-análise, replicação, meta-análise, etc.).

Hyman (1972) enfatiza que as habilidades para o reuso de dados devem ser construídas por meio de intensa formação metodológica. Hyman concentra-se principalmente sobre a importância do conhecimento estatístico, considerando que os seus argumentos são centrados na análise secundária de dados quantitativos. Corti e Bishop (2005), por outro lado, defendem a necessidade de programas que desenvolvam técnicas e habilidades para a reutilização de dados qualitativos e a importância de programas de formação para pesquisadores, tanto para

construir a consciência no que tange à oportunidade geral para o reuso de dados, quanto para o desenvolvimento de competências específicas para a prática do reuso. Da mesma forma, Kriesberg et al. (2013) articulam sobre a importância destes treinamentos especialmente para pesquisadores novatos.

Não obstante ficou patente nos discursos dos entrevistados que o reuso de dados compreende um conjunto de habilidades específicas, que requer prática e instrução. Em muitos casos, a participação em treinamentos foi determinante para o reconhecimento das possibilidades de reuso e para as experiências de reuso efetivo dos entrevistados.

4.6 INFLUÊNCIA SOCIAL

A sexta e última variável teórica do estudo corresponde aos elementos do ambiente social dos cientistas que podem influenciar, tanto suas intenções, quanto suas experiências de reuso de dados. A pesquisa revelou dois aspectos que foram classificados na dimensão da influência social: a “receptividade disciplinar” e o “incentivo dos pares”.

Há uma suposição geral de que alguns campos de estudo e disciplinas sejam mais receptivos ao reuso de dados (BORGMAN, 2007; FANIEL; JACOBSEN, 2010, THESEN; PATTERNSON, 2011) e acolham melhor essa prática. Mesmo que este estudo exploratório de pequena escala não tenha tido por objetivo de realizar qualquer tipo de análise comparativa entre diferentes disciplinas das ciências sociais, ficou patente na fala dos entrevistados a importância da receptividade disciplinar.

Além das normas disciplinares, o incentivo proveniente dos pares – nem sempre pertencentes à mesma disciplina –, foi identificado com um importante motivador para o reuso de dados, principalmente quando a possibilidade de reuso é apresentada como uma opção por mentores ou pesquisadores mais experientes aos seus pupilos.

Outra pessoa no meu laboratório estava utilizando os dados do repositório de Harvard (...), assim meu professor me apresentou este diretório de dados e este rico conjunto de dados (Nathan).

Esta noção de incentivo coaduna com os argumentos de Kriesberg et al. (2013) sobre o papel dos investigadores seniores em modelar e incutir a prática de reuso de dados em pesquisadores iniciantes.

5 CONSIDERAÇÕES FINAIS

Os resultados obtidos por meio do estudo exploratório sustentam o pressuposto inicial da pesquisa de que os cientistas ponderam uma combinação de fatores além das motivações frugais ao considerarem o reuso de dados em suas pesquisas. Como visto, a parcimônia representa apenas um dos fatores dentre os 25 fatores identificados nas seis diferentes dimensões passíveis de influenciar a intenção e o comportamento de cientistas quanto ao reuso de dados. O reuso de dados, portanto, não se configura somente como uma estratégia de atalho, mais rápida e menos onerosa para a produção científica. Pode também ser uma alternativa para a expansão e o avanço da ciência, seja pela exploração de diferentes vieses em um mesmo conjunto de dados, confirmação ou refutação de análises anteriores, baseados em dados que, na perspectiva dos reutilizadores foram de certo modo, minimamente avaliados e endossados pelos repositórios que os disponibilizam.

Além dos benefícios associados ao reuso, por meio das entrevistas e dos diferentes relatos de experiências dos participantes, ficou patente que cientistas consideram diferentes condições tanto antes, quanto durante o processo de reuso de dados coletados por outros pesquisadores e disponibilizados em repositórios para acesso aberto. Estas condições incluem potenciais riscos associados ao reuso de dados que podem desmotivá-los a optar por essa abordagem de pesquisa, os diferentes esforços associados ao trabalho com dados coletados por outros pesquisadores, o potencial de reusabilidade dos dados, a existência de condições de infraestrutura e pessoal que facilitem esse processo, bem como, a receptividade de seus pares e de suas disciplinas quanto à esta prática.

Muito embora os resultados sejam baseados em apenas cerca de uma dezena de entrevistas eles oferecem um importante plano conceitual preliminar para pesquisas que busquem investigar questões relativas ao comportamento de reuso de dados de pesquisa no contexto da comunicação científica e da sociologia da ciência. As dimensões (variáveis teóricas)

e fatores aqui apresentados abrem possibilidades para estudos futuros com vistas a propor e testar modelos sobre o comportamento de reuso de dados científicos. Compreender como estes e outros fatores podem afetar as intenções e práticas de reuso de dados pode contribuir também para a definição de políticas de incentivo ao compartilhamento e ao reuso de dados, e para o estabelecimento de repositórios de dados melhor alinhados às expectativas da comunidade científica.

REFERÊNCIAS

BEHBOUDI, M. *et al.* **Human Knowledge Intermediaries for Successful Knowledge Re-use and Sharing**. New York, 2008. Paper presented at the 5th International Conference on Intellectual Capital, Knowledge Management & Organizational Learning (ICICKM).

BOH, W. F. Re-use of knowledge assets from repositories: a mixed methods study. **Information & Management**, Harrisburg, v. 45, n. 6, p. 365-375, 2008.

BORGMAN, C. L. **Scholarship in the digital age**. Cambridge: The MIT Press, 2007.

BORGMAN, C. L. *et al.* **Data, Data Use, and Scientific Inquiry**: two case studies of data practices. Los Angeles: ACM Press, 2012.

BOSLAUGH, S. **An introduction to secondary data analysis**. Secondary Data Sources for Public Health: a practical guide. Cambridge: Cambridge University, 2007. Disponível em: < http://www.academia.edu/1630213/An_introduction_to_secondary_data_analysis>. Acesso em: 22 jun. 2016.

CARLSON, S. *et al.* What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. **Journal of Computer-Mediated Communication**, Malden, v.12, n. 2, p. 635-651, 2007.

CASTLE, J. E. Maximizing research opportunities: Secondary data analysis. **Journal of Neuroscience Nursing**, Chicago, v. 35, n. 5, p. 287-290, 2003.

CHAO, T. C. Disciplinary reach: Investigating the impact of data set re-use in the earth sciences. **Proceedings of the American Society for Information Science and Technology**, New York, v. 48, n. 1, p. 1-8, 2011.

CHAO, T. C. **Exploring the rhythms of scientific data use**. 2012. Paper presented at the Proceedings of the iConference. 2012. doi:10.1145/2132176.2132193

CORTI, L. *et al.* Strategies in teaching secondary analysis of qualitative data. **Forum: Qualitative Social Research**, Berlin, v. 6, n. 1, 2005.

CORTI, L.; THOMPSON, P. Secondary analysis of archived data. In: _____. **Qualitative Research Practice**. Thousand Oaks: Sage Publications, 2004. p. 297-314.

Tendências da Pesquisa Brasileira em Ciência da Informação, v.9, n.2, set./dez. 2016.

CURTY, R. G. **Beyond data thrifting: an investigation of factors influencing research data re-use in the social sciences**. 2015. Thesis (Doctoral) – Syracuse University, Syracuse. 2015. Disponível em: <<http://surface.syr.edu/etd/266/>>. Acesso em: 18 jul. 2016.

DALE, A. Secondary Analysis of Quantitative Data. In: LEWIS-BECK, M. A. (Ed.). **The Sage Encyclopedia of Social Science Research Methods**. Thousand Oaks: Sage Publications, 2004. v. 1.

DARBY, R. *et al.* **Enabling scientific data sharing and re-use**. 2012. Paper presented at the E-Science (e-Science), IEEE International Conference on Intelligent Systems. 8. 2016. Sofia, Bulgaria: IEEE, 2016. p. 1-8.

DAVID, M. The science of data sharing: documentation. In: SIEBE, J. E. (Ed.). **Sharing social science data: advantages and challenges**. Thousand Oaks: Sage Publications, 1991. p. 91-115.

DAVIS, H. M.; VICKERY, J. N. Datasets, a shift in the currency of scholarly communication: Implications for library collections and acquisitions. **Serials Review**, Greenwich, v. 33, n. 1, p. 26-32, 2007.

FAHS, P. S. *et al.* M. A call for replication. **Journal of Nursing Scholarship**, Indianapolis, v. 35, n. 1, p. 67-72, 2003.

FANIEL, I. *et al.* **An Exploratory Study of the Factors Associated with Successful Re-use of Others Knowledge**. 2002. Paper presented at the Americas Conference on Information Systems (AMCIS).

FANIEL, I. M. *et al.* Re-using scientific data: how earthquake engineering researchers assess the re-usability of colleagues' data. **Computer Supported Cooperative Work**, Dordrecht, v. 19, n. 3, p. 355-375, 2010.

FANIEL, I. M. *et al.* Beyond the data deluge: a research agenda for large-scale data sharing and re-use. **International Journal of Digital Curation**, Edinburgh, v. 6, n. 1, p. 58-69, 2011.

FANIEL, I. M. *et al.* **A Comparative Study of Data Re-use Among Quantitative Social Scientists and Archaeologists**. 2013. Paper presented at the iConference Fort Worth.

FANIEL, I. M. **Data Re-use and Sensemaking among Novice Social Scientists**. Paper presented at the ASIST 2012 Annual Meeting, Baltimore, MD. 2012.

FANIEL, I. M. *et al.* **Can Quantitative Social Scientists Get Data Re-use Satisfaction?** Baltimore, 2013. Panel presented at the Research Data Access and Preservation Summit.

FIENBERG, S. *et al.* (Ed.). **Sharing research data**. Washington, DC: National Academy Press, 1985.

HEATON, J. **Reworking Qualitative Data**. Thousand Oaks: Sage Publications, 2004.

HINDS, P. S. *et al.* The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. **Qualitative Health Research**, Thousand Oaks, v. 7, n. 3, p. 408-424, 1997.

HOWARD, T. *et al.* **Opportunities for and Barriers to Engineering Research Data Re-use**. Bath: University of Bath, 2010. Disponível em: <<http://opus.bath.ac.uk/21166/1/erim3res100805tjh10.pdf>>. Acesso em: 12 jul. 2015.

HYMAN, H. H. **Secondary Analysis of Sample Surveys**: principles, Procedures and Potentialities. New York: John Wiley & Sons, 1972.

KANKANHALLI, A. *et al.* Knowledge re-use through electronic repositories: a study in the context of customer service support. **Information & Management**, Harrisburg, v. 48, n. 2, p. 106-113, 2011.

KELDER, J. A. Using Someone Else's Data: Problems, Pragmatics and Provisions. **Forum: Qualitative Social Research**, Berlin, v. 6, n. 1, Jan. 2005. Disponível em: <<http://www.qualitative-research.net/index.php/fqs/article/view/501/1078>>. Acesso em: 12 jul. 2015.

KIECOLT, K. J. *et al.* **Secondary analysis of survey data**. Beverly Hills: Sage Publications, 1985.

KLEIN, G. *et al.* **A data-frame theory of sensemaking**. New York: Lawrence Erlbaum Associates, 2007.

KRIESBERG, A. *et al.* **The Role of Data Re-use in the Apprenticeship Process**. Montreal, 2013. Paper presented at the 76th Annual Meeting of the Association for Information Science and Technology,

LAW, M. Reduce, Re-use, Recycle: Issues in the Secondary use of Research Data. **IASSIST Quarterly**, Odense, v. 29, n. 1, p. 5-10, Spring 2005.

MARCIAL, L. H. Scientific data repositories on the Web: an initial survey. **Journal of the American Society for Information Science and Technology**, New York, v. 61, n. 10, p. 2029-2048, ag. 2010.

MARKUS, M. L. Toward a theory of knowledge re-use: Types of knowledge re-use situations and factors in re-use success. **Journal of Management Information Systems**, Armonk, v. 18, n. 1, p. 57-94, 2001.

MARTIN, P. More replication studies needed. **Applied Nursing Research**, Philadelphia, v. 8, n. 2, p. 102, 1995.

MORAES, R. Análise de conteúdo. **Revista Educação**, Porto Alegre, v. 22, n. 37, p. 7-32, 1999.

NIU, J. *et al.* **Documentation Evaluation Model for social science data**: an empirical test. Michigan, 2009. Proceedings of DigCCurr 2009 Digital Curation: Practice, Promise and Prospects, 125.

NIU, J. Overcoming inadequate documentation. **Proceedings of the American Society for Information Science and Technology**, Michigan, v. 46, n. 1, p. 1-14, 2009a.

Tendências da Pesquisa Brasileira em Ciência da Informação, v.9, n.2, set./dez. 2016.

NIU, J. **Perceived Documentation Quality of Social Science Data**. 2009. Thesis. (Doctor of Philosophy Dissertation)-University of Michigan, Ann Arbor, 2009b. Disponível em: <http://deepblue.lib.umich.edu/bitstream/handle/2027.42/63871/niuif_1.pdf?sequence=1>. Acesso em: 10 nov. 2015.

PALMER, C. L. *et al.* The analytic potential of scientific data: understanding re-usevalue. **Proceedings of the American Society for Information Science and Technology**, Malden, v. 48, n. 1, p. 1-10, 2011.

PIGOTT, D.; HOBBS, V.; GAMMACK, J. An approach to managing repurposing of digitized knowledge assets. **Australasian Journal of Information Systems**, Malden, v. 9, n. 1, p. 92-103, 2001.

PIWOWAR, H. *et al.* **Data re-use and the open data citation advantage**. 2013. 1(e175). Disponível em: <<https://peerj.com/articles/175/>>. Acesso em: 12 jun. 2016.

PIWOWAR, H. A. Proposed Foundations for Evaluating Data Sharing and Re-use in the Biomedical Literature. **Bulletin of the IEEE Technical Committee on Digital Libraries**, Pittsburgh, v. 4, n. 2, Fall 2008. Disponível em: <<http://www.ieee-tcdl.org/Bulletin/v4n2/piwowar/piwowar.html>>. Acesso em: 12 jun. 2016.

PIWOWAR, H. A. **Foundational studies for measuring the impact, prevalence, and patterns of publicly haring bio medical research data**. 2010. Thesis (Doctor of Philosophy Dissertation)-University of Pittsburgh, Pittsburgh, 2010a.

PIWOWAR, H. A. A method to track data setre-use in biomedicine: filtered GEO accession numbers in Pub Med Central. **Proceedings of the American Society for Information Science and Technology**, Malden, v. 47, n. 1, p. 1-2, 2010b. doi: 10.1002/meet.14504701450

SANDS, A. *et al.* Follow the Data: How Astronomers Use and Re-use Data. Paper presented at the ASIST 2012. **Proceedings of the American Society for Information Science and Technology (ASIST)**, Baltimore, p. 28-31, Oct. 2012.

TENOPIR, C. *et al.* Data sharing by scientists: practices and perceptions. **PLoS ONE**, San Francisco, v. 6, n. 6, p. e21101, 2011. doi: 10.1371/journal.pone.0021101.

THESSSEN, A. E. *et al.* Data issues in the life sciences. **ZooKeys**, Sofia, v. 15, n.150, nov. 2011. p. 15-51. doi: 10.3897/zookeys.150.1766

WALLIS, J. C. *et al.* If We Share Data, Will Anyone Use Them? Data Sharing and Re-use in the Long Tail of Science and Technology. **PloS One**, San Francisco, v. 8, n. 7, p. e67332, jul. 2013.

ZIMMERMAN, A. Not by metadata alone: the use of diverse forms of knowledge to locate data for re-use. **International on Digital Libraries**, Heidelberg, v. 7, n. 1/2, p. 5-16, 2007.

ZIMMERMAN, A. S. **Data sharing and secondary use of scientific data**: experiences of ecologists. 2003. Thesis (Doctor of Philosophy in Information and Library Studies Dissertation)-The University of Michigan, Michigan, 2003.

ZIMMERMAN, A. S. New Knowledge from Old Data: The Role of Standards in the Sharing and Re-use of Ecological Data. **Science, Technology & Human Values**, Ann Arbor, v. 33, n. 5, p. 631-652, 2008.

XIA, J. *et al.* (2012). A Review of Open Access Self-Archiving Mandate Policies. **Libraries and the Academy**, Baltimore, v. 12, n. 1, p. 85-102.

APÊNDICE A - SÍNTESE DA ANÁLISE DAS ENTREVISTAS.

Dimensões/ Fatores	N	R	Descrição	Literatura Relacionada
Benefícios Percebidos				
Expansão do Conhecimento	9	17	Dados podem ser reutilizados para atender a diferentes propósitos, bem como para replicação e validação de pesquisas com vistas à expandir os horizontes do conhecimento científico	Darby <i>et al.</i> (2012) Hyman (1972) Kankanhalli, Lee e Lim (2011)
Parcimônia	13	63	Maneiras de contornar problemas de coleta de dados associados com o tempo e custo (dinheiro) para minimizar os esforços duplicados	Castle (2003) Hyman (1972) Kiecolt and Nathan (1985) Law (2005)
Pré-Endossamento	5	6	Estima-se que dados disponíveis para reuso são considerados de alguma forma credível e fiável, caso contrário não seriam compartilhados e disponibilizados publicamente	N/A
Riscos Percebidos				
Desvalorização	2	2	Pesquisas que reutilizam dados podem ser consideradas menos autênticas e de menor impacto, gerando menor prestígio e reconhecimento ao reutilizador	Goodwin (2012) Fahs, Morgan e Kalman (2003) Martin (1995)
Infração de Códigos Éticos	1	1	Possíveis incautos de violação de aspectos de confidencialidade, direitos de autor e de proteção de dados	Grinyer (2009) Heaton (2004) Law (2005)
Deslize	3	7	Má interpretação, associação incorreta, ou uso indevido que possa ocorrer durante o reuso de dados de outros pesquisadores.	Corti e Thompson (2004) Kuula (2010) Tenopir <i>et al.</i> (2011)
Vulnerabilidade a Erros	8	10	A susceptibilidade a erros dada à dificuldade de identificar possíveis problemas em dados coletados/produzidos por outros pesquisadores.	Castle (2003) Hyman (1972) Kiecolt e Nathan (1985)
Esforços Percebidos				
Inovar com dados antigos	5	7	Necessidade de identificar contribuições originais a partir de dados de segunda mão e explorar diferentes questões ainda não exploradas ou negligenciadas pelos pesquisadores principais, bem como por outros reutilizadores	Zimmerman (2008)
Acesso à dados	3	7	O esforço associado à obtenção de acesso e recuperação de dados para potencial reuso	Faniel <i>et al.</i> (2013) Heaton (2004) Tenopir <i>et al.</i> (2011)
Descoberta dos Dados	5	6	O esforço associado com a encontrabilidade de dados e a identificação de conjuntos de dados relevantes e potencialmente reutilizáveis	Darby <i>et al.</i> (2012) Faniel e Majchrzak (2002)
Lidar com Incompatibilidades	12	75	Necessidade de adaptação dos dados produzidos a partir de questões de pesquisa e/ou hipóteses específicas, utilizando instrumentos ou técnicas de coleta de dados particulares, em um contexto particular e calendário, e com variáveis específicas. Esforço associado à renúncia das ideias iniciais e reformulação do estudo a fim de acomodar os dados existentes	Dale (2004) Heaton (2004) Kiecolt and Nathan (1985) Klein <i>et al.</i> (2007)

Preparação para o reuso	9	35	Preparo e manipulação pré-reuso, incluindo: seleção, triagem e limpeza, recodificação e formatação dos dados.	Faniel <i>et al.</i> (2013) Heaton (2004)
Compreensão do Estudo Original	8	29	Necessidade de atribuir sentido aos dados e compreender o estudo original em sua totalidade	Boh (2008) Faniel e Jacobsen (2010)
Reusabilidade dos Dados				
Qualidade da Documentação	6	24	O quão completa, e clara é a documentação complementar aos dados que descreve os fundamentos, a metodologia do estudo, e os dados.	David (1991) Faniel <i>et al.</i> (2013) Niu e Hedstrom (2009) Niu (2009ab) Zimmerman (2003; 2007; 2008)
Adaptabilidade dos Dados	9	19	O quão o assunto, o nível de análise, e o tipo de dados são compatíveis com a finalidade/propósito de reuso	Faniel <i>et al.</i> (2013) Hinds, Vogel e Clarke-Steffen (1997) Palmer, Weber e Cragin (2011)
Credibilidade dos Produtores de Dados	7	9	O quão confiáveis os produtores/coletores de dados (instituições, autores individuais ou grupos) são	Boh (2008) Carlson e Anderson (2007) Darby <i>et al.</i> (2012) Faniel <i>et al.</i> (2013)
Qualidade dos Dados	7	14	Consistência e completude dos dados	Faniel <i>et al.</i> (2013) Hinds, Vogel e Clarke-Steffen (1997)
Rigor do Estudo Original	3	8	Qualidade do design metodológico e da execução da pesquisa	Hinds, Vogel e Clarke-Steffen (1997)
Condições Facilitadoras				
Existência de Documentação de Dados	6	19	A disponibilidade de documentação de dados de forma detalhada e completa	David (1991); Faniel e Majchrzak (2002); Markus (2001); Pigott, Hobs e Gammack (2001); Niu e Hedstrom (2009); Niu (2009ab); Zimmerman (2003; 2007; 2008)
Existência de Repositórios de Dados	5	9	A existência de repositórios e sua capacidade de organizar, preservar, e facilitar o acesso a dados com potencial de reuso	Markus (2001) Palmer, Weber e Cragin (2012) Tenopir, et al (2011)
Alcance dos Investigadores Primários	5	10	A possibilidade de comunicação direta com os investigadores primários como mecanismo para a obtenção de informações adicionais sobre os dados e o estudo	Boh (2008) Faniel and Zimmerman (2011)
Suporte e Assistência	5	16	Apoio institucional e/ou da equipe de assessoramento do repositório de dados	Behboudi e Hart (2008) Faniel <i>et al.</i> (2013) Markus (2001)
Treinamento e Formação	8	21	Importância da formação e programas de treinamento em análise secundária e técnicas metodológicas orientadas ao reuso	Corti e Bishop (2005) Hyman (1972) Kriesberg <i>et al.</i> (2013)
Influência Social				

Receptividade Disciplinar	10	12	Tradição disciplinar ou acolhimento à prática de reuso de dados. Algumas disciplinas são mais propensas ao reuso de dados.	Borgman (2007) Faniel e Jacobsen (2010) Thessen e Patterson (2011)
Incentivo dos Pares	4	4	A aceitação ou prática habitual de reuso entre os pares. O encorajamento advindo de pesquisadores mais experientes com a prática de reuso de dados.	Kriesberg <i>et al.</i> (2013)

Fonte: Dados da Pesquisa.