

PUBLICAÇÕES AMPLIADAS: aspectos da integração de dados de pesquisa

*Daiane Barrili dos Santos**

*Moisés Rockembach***

RESUMO

Este artigo apresenta resultados de pesquisa, que teve como objetivo principal propor um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas. O modelo de análise deste estudo foi composto por três fases, iniciando pela revisão de literatura da área, passando pela identificação de repositórios de dados do diretório re3data.org que utilizam ligações baseadas no modelo de publicação ampliada e, por fim, a inspeção das interfaces dos repositórios selecionados. Verificou-se uma necessidade crescente na adoção de uma infraestrutura compatível com as publicações ampliadas, tornando-se importante a utilização de requisitos para as interfaces dos repositórios de dados de pesquisa. O cruzamento de informações realizado neste estudo permitiu um maior entendimento do quanto é relevante o uso de cada um dos requisitos estudados, como também foi possível compreender as consequências caso estes requisitos não estejam contemplados nas interfaces de repositórios. A partir da investigação e do aprofundamento sobre o tema, foi possível ampliar o conhecimento e detectar os elementos apropriados para propor um panorama de integração entre um repositório de dados de pesquisas e publicações científicas baseando-se em um modelo de publicação ampliada, como também destacar as vantagens da utilização deste modelo.

Palavras-chave: Publicações ampliadas. Repositórios de dados de pesquisa. Requisitos para publicações ampliadas.

* Mestre em Comunicação e Informação pelo Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul, Brasil. Doutoranda em Comunicação e Informação pelo Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul, Brasil.
E-mail: dai.b.santos@hotmail.com.

** Doutor em Informação e Comunicação em Plataformas Digitais pela Universidade do Porto, Portugal. Docente permanente do Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul, Brasil.
E-mail: moises.rockembach@gmail.com.

I INTRODUÇÃO

A Comunicação Científica vem sofrendo transformações com o advento da web e os pesquisadores têm se apropriado das tecnologias oferecidas para disseminar as produções científicas através de diversas alternativas de publicações. Da mesma forma que há o interesse em dar o acesso livre às publicações científicas por meio de repositórios digitais, atualmente também existe vantagem em dar acesso livre aos dados provenientes da pesquisa científica.

Na comunidade científica utilizam-se diferentes tipos de publicações para a disseminação de trabalhos científicos, os quais podem estar disponíveis em ambientes digitais. A divulgação de trabalhos científicos e o desenvolvimento da ciência estão diretamente ligados à busca de melhoria contínua das plataformas de disseminação científica. Deste modo, o aperfeiçoamento dessas tecnologias possibilita o processo de disponibilização de materiais digitais por meio de ferramentas que possuem a finalidade de organizar a informação e de disseminar o uso do conhecimento.

Essas inovações da comunicação científica também estão relacionadas à Ciência Aberta (*Open Science*). Conforme Sayão e Sales (2014), a ideia de Ciência Aberta tem muitos significados, porém o mais convincente é o que reconhece que o conhecimento científico é um patrimônio da humanidade e, portanto, precisa estar disponível livremente para que as pessoas possam usá-lo, reusá-lo e distribuí-lo. Por esta razão, a Ciência Aberta pode proporcionar maior dinamismo às atividades que envolvem a ciência, a tecnologia e a inovação. Albagli, Maciel e Abdo (2015) enfatizam que o movimento pela Ciência Aberta se insere nas novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento e da cultura.

Nesse contexto, a internet torna-se uma aliada da pesquisa e da comunicação científica devido ao seu potencial de ampliar o alcance de resultados de pesquisas e publicações científicas. Com a popularização da internet, surgiram diversas melhorias nos canais de comunicação científica como, por exemplo, as publicações on-line que revolucionaram as possibilidades de descoberta e de acesso a recursos informacionais acadêmicos e científicos aumentando, desta forma, a amplitude e a facilidade de acesso a todo tipo de material, até mesmo conjuntos de dados científicos (SALES; SAYÃO; SOUZA, 2013). Na Web está disponível uma imensurável quantidade destes dados e na tentativa de estruturar e tornar acessível estes dados surge a Web Semântica que “[...] é o uso de uma linguagem formal para representar o conhecimento conceitual de um determinado domínio de conhecimento presente na Web de forma estruturada e compreensível.” (SOUZA, 2013, p. 1).

Com o aprimoramento de tecnologias que possibilitam a armazenagem e o compartilhamento de informações, iniciativas de acesso aberto vêm alterando as formas de disseminação e uso dessas informações. O movimento de Acesso Aberto (*Open Access*) ou acesso livre, que surgiu com a Declaração de Budapest em 2002, só aconteceu a partir da contribuição da Iniciativa dos Arquivos Abertos - *Open Archives Initiative* (OAI). No entanto, a Declaração de Berlin (2003) ampliou o propósito do que se entendia por acesso aberto. Conforme a Declaração de Berlin (2003), as contribuições de acesso livre

incluem resultados de pesquisas científicas originais, ou seja, os dados não processados e as fontes originais. Desta forma, considera-se que o conceito de acesso aberto não está somente relacionado a publicações acadêmicas tradicionais como também pode abranger coleções de dados de pesquisa.

Por sua vez, as tecnologias digitais têm transformado a forma como esses conjuntos de dados eram caracterizados. Após o término de pesquisas e publicações, os dados eram muitas vezes descartados ou armazenados em servidores sem a devida gestão (SAYÃO; SALES, 2014). Nesse seguimento, outros movimentos e iniciativas surgem nos fazendo repensar os modelos atuais de disseminação de resultados e a compreender a otimização de pesquisas.

Alguns modelos de publicações científicas estão sendo desenvolvidos para que haja um maior compartilhamento e reuso de informações, um destes modelos é a publicação ampliada. Nesta pesquisa é apresentada esta nova forma de publicação - *Publicação Ampliada*, do inglês *Enhanced Publication* - que depende das possibilidades de ligações da web. Esse tipo de publicação proporciona a colaboração entre pesquisadores, a conectividade, o compartilhamento de dados e as convergências de linguagens. Nesse sentido, acredita-se que um conjunto de dados abertos de pesquisa, disponíveis em um repositório digital, possibilita o reuso, o aumento do índice de citações de autores, potencializa a troca de conhecimento e evita que futuras pesquisas sobre um determinado tema se repitam, evitando desperdícios financeiros e de tempo.

Iniciativas referentes à abertura de dados de pesquisa são apoiadas por uma ampla gama de políticas. Agências internacionais de financiamento à pesquisa científica como, por exemplo, a *Organization for Economic Cooperation and Development* (OECD) e revistas científicas como a *Science and Nature* recomendam que os dados de pesquisas científicas estejam em acesso aberto ao público. Tornar dados disponíveis publicamente já faz parte da cultura em algumas áreas como a Física, a Astronomia, a Biologia, entre outras. Desta forma há o interesse das instituições em aderir às recomendações e desenvolver planos de gestão de dados de pesquisa que estabeleçam de forma clara todas as ações e requisitos para a abertura dos dados

de pesquisa em seus repositórios. Quanto as agências internacionais de financiamento à pesquisa, além da OECD, podemos citar a *National Science Foundation* (NSF), o *National Institute of Health* (Estados Unidos) e a *The Royal Society* (Reino Unido), que também começaram a estabelecer políticas para disponibilizar esses dados de pesquisas. Estas agências tem recomendado as instituições que divulguem junto às publicações os dados provenientes de pesquisas científicas por meio de repositórios digitais. Além disso, recomendam a utilização de planos de gestão de dados e oferecem suporte aos pesquisadores para a construção e desenvolvimento de repositórios de dados de pesquisa.

A NSF é uma agência federal independente criada pelo Congresso Americano que promove o progresso da ciência e apoia a pesquisa. Como parte da proposta de financiamentos, a agência exige um plano de gerenciamento de dados¹ (ICPSR, 2017). Da mesma forma que a NSF, acredita-se na tendência de que a maioria das agências financiadoras irá exigir algum tipo de plano de gerenciamento de dados.

No Brasil, até o momento, não existem agências de fomento que tenham desenvolvido iniciativas para incentivar a disponibilização e compartilhamento de dados de pesquisa. No entanto, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), estende a visão sobre o tema e reconhece a importância do acesso aberto a dados de pesquisa disponibilizando um repositório de dados de pesquisas usando a ferramenta *Dataverse*.

O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O *Dataverse* é um grande repositório aberto de dados de todas as disciplinas, criado pelo *Institute for Quantitative Social Science* da Universidade de Harvard. O repositório *Dataverse* do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana. (IBICT, 2017).

¹ The National Institutes of Health (NIH) now requires a data sharing plan for large projects, and in 2011 the National Science Foundation (NSF) began to require a data management plan as part of every grant application. (ICPSR, 2017).

Percebe-se que essas inovações já estão presentes em outros países e que no Brasil ainda é bastante recente. Devido a este motivo, organizações que possuem a incumbência de promover o desenvolvimento tecnológico e apoiar pesquisas no Brasil como, por exemplo, a Rede Nacional de Ensino e Pesquisa (RNP), estão criando projetos relacionados ao desenvolvimento de repositórios de dados de pesquisa para a publicação e partilha desses dados. A RNP organiza a integração global e a colaboração apoiada em tecnologias de informação e comunicação para a geração do conhecimento e a excelência da educação e da pesquisa (RNP, [201-]). Desde 2002 é uma Organização Social (OS) vinculada ao Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) e mantida por esse em conjunto com os ministérios da Educação (MEC), Cultura (MinC), Saúde (MS) e Defesa (MD), que participam do Programa Interministerial da RNP (PI-RNP). Grandes instituições de ensino estão se empenhando para promover o acesso aberto aos dados das pesquisas científicas e desenvolvendo diretrizes para a implementação de repositórios e planos de gestão de dados.

Entre as ações envolvidas no desenvolvimento de um repositório de dados de pesquisa está a publicação ampliada, com o propósito de conectar os dados das pesquisas com suas respectivas publicações. Observou-se que temas relacionados aos dados abertos ainda carecem de estudos que apresentem de forma clara como esses dados podem ser usados, reutilizados e compartilhados, necessitando rever o modelo atual de disseminação de resultados de pesquisas.

2 PROCEDIMENTOS METODOLÓGICOS

Para que fosse possível verificar um modelo de publicação ampliada, foi necessário identificar na literatura referente à área os principais elementos utilizados para a ligação dos dados às publicações científicas. Da mesma forma, foi necessário identificar nas interfaces dos repositórios esses elementos para que fosse possível realizar comparações entre as informações coletadas na literatura e os dados coletados nas interfaces dos repositórios. A verificação de requisitos presentes nas interfaces possibilitou a

identificação de aspectos relacionados à publicação ampliada utilizado entre os repositórios.

Ainda na execução da pesquisa, foi elaborado um protocolo de revisão sistemática da literatura, que é uma metodologia de estudo que visa estabelecer as etapas devidamente delineadas e definidas em todos os momentos, ou seja, deve esclarecer o modo como serão selecionadas e apuradas as fontes, de maneira que as conclusões a extrair sobre os assuntos em estudo possam ser cientificamente consistentes. É importante esclarecer a diferença entre uma pesquisa bibliográfica e uma revisão sistemática. De acordo com Kitchenham (2004, p. 3), o processo para a condução de revisões sistemáticas envolve três etapas: planejamento da revisão, condução da revisão e publicação (relatório) dos resultados.

No decorrer do planejamento da revisão, o protocolo de revisão foi definido. Durante a condução da revisão, as fontes para a revisão sistemática foram selecionadas e os estudos primários foram identificados, selecionados e avaliados de acordo com critérios de exclusão, inclusão e de qualidades. Os critérios de inclusão e exclusão servem para indicar por quais motivos um estudo foi incluído ou excluído da revisão ou porque são considerados relevantes ou não para o estudo. Com os critérios de qualidade foi possível avaliar aspectos metodológicos dos estudos, buscando o aumento da confiabilidade nos resultados que serão obtidos. Após a seleção dos estudos, os dados foram extraídos e sintetizados para serem estabelecidas as categorias para análise.

Para auxiliar nos registros durante a etapa da condução da revisão, foram utilizados formulários. Através do método de revisão sistemática, foram identificados estudos sobre o tema em questão para extrair os principais requisitos de uma publicação ampliada e averiguar a aplicabilidade. Para a realização da análise e comparações nos repositórios de acordo com os requisitos recuperados durante a revisão sistemática, foram estabelecidas as categorias para a análise. As categorias são classes que reúnem um grupo de elementos sob um título genérico, agrupamento esse efetuado em razão dos caracteres comuns destes elementos (BARDIN, [19- -]).

Figura 1: Categorias para análise

| Categorias para análise |
|---|
| Recursos WEB - URI |
| Identificador persistente |
| Vínculo de objetos digitais as publicações |
| Registro das versões das publicações e dados |
| Registro das propriedades básicas (publicação e recursos adicionados) |
| Vocabulário controlado |
| Atributos - descrição |
| Registro de autoria da publicação e dados de pesquisa |
| Preservação |
| Registro das relações que fazem parte da publicação ampliada |
| Modelo OAI-ORE |
| Open access |
| Padrões web semântica |
| Relacionamento semântico |
| Upload de objetos digitais |
| Metadados |

Fonte: Elaborado pelos autores

A partir das categorias definidas, um *checklist* foi elaborado com base em critérios e elementos considerados essenciais no desenvolvimento de um modelo de publicação ampliada. O *checklist* auxiliou nas observações dos repositórios e identificação dos requisitos presentes em suas interfaces e a comparar com os requisitos estabelecidos neste estudo, através da literatura recuperada na revisão sistemática.

Para obter os registros e fazer a seleção dos repositórios de dados de pesquisa, foi utilizado o diretório re3data.org (*Registry of Research Data Repositories*) que caracteriza-se por ser uma ferramenta de busca de repositórios de dados. No período de 2012 a 2016, mais de 1.500 repositórios foram indexados.

O re3data.org é um registro global de repositórios de dados de pesquisa que cobre repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Ele apresenta repositórios para o armazenamento permanente e acesso de conjuntos de dados para

pesquisadores, organismos de financiamento, editores e instituições acadêmicas. Ele promove uma cultura de partilha, maior acesso e melhor visibilidade dos dados da investigação. Para o registro das observações durante a etapa de análise dos repositórios, foi utilizado um *checklist* de validação de requisitos. Para realizar a inspeção, foi utilizado como roteiro de observação um *checklist* com os requisitos essenciais para a implementação de um modelo de publicações ampliadas, conforme a indicação da literatura da área.

3 PUBLICAÇÃO AMPLIADA

As publicações científicas têm como propósito disseminar a pesquisa para a comunidade de modo que permita que outros pesquisadores possam avaliá-la, utilizá-la e reutilizá-la sob outro ponto de vista. Através de uma publicação científica, é possível ter o conhecimento dos resultados de uma pesquisa. A discussão proposta nesta seção é aprofundar o conhecimento dos modelos de publicações científicas em um momento em que o modelo tradicional tem sido crescentemente questionado, apontando para a necessidade de discussões sobre novos modelos. De acordo com Nassi-Calò (2016):

O artigo científico, seu formato e meios de publicação, disseminação e compartilhamento passou por significativas alterações desde o surgimento dos primeiros periódicos científicos no século XVII. A Internet, nos anos 1990, mudou drasticamente o paradigma da comunicação da ciência, evento comparável apenas à invenção da imprensa por Gutenberg em 1440, que possibilitou a disseminação dos artigos e periódicos para outras instâncias, além da academia. (NASSI-CALÒ, 2016).

Nota-se que a impressão em papel por meio de tipos móveis trouxe, além da possibilidade de reprodução das publicações em larga escala, uma maneira específica de ler o texto. Conforme Castedo (2007), sempre que há a criação de um novo suporte, cria-se também um novo hábito de leitura. Em meados do século XIX, Paul Otlet, “pai da Documentação e da internet”,

já criticava o modelo tradicional de publicação científica e influenciou pesquisadores que o sucederam no desenvolvimento de ferramentas para o acesso à informação. Em sua obra *Traité de Documentation: le livre sur le livre: théorie et pratique*, publicada em 1934, o autor trata de muitas ideias, hoje adotadas, sobre a organização de redes de cooperação para tratamento e troca de informações documentadas. Nesta obra, Paul Otlet apresenta os produtos do desenvolvimento tecnológico de sua época e suas aplicações à documentação, propondo uma rede universal de informação e documentação. (SANTOS, 2007). Na década de 1980, o desenvolvimento das tecnologias da informação e comunicação alcançou a etapa em que podia começar a competir com a impressão em papel como meio universal para difundir informações científicas. Nos últimos anos, portanto, passou a ser importante examinar a possibilidade de se transferir informações científicas do meio impresso para o meio eletrônico. (MEADOWS, 1999).

Com o advento da Web, em 1990, diversas possibilidades de melhorias nos canais de comunicação científica surgiram, o que ocasionou a necessidade de novos modelos de publicações. Neste contexto, é importante a reflexão sobre as mudanças que a comunicação de informações vem sofrendo na última década. As tecnologias da informação possibilitaram o acesso às publicações científicas eletrônicas, facilitando a disseminação da pesquisa, permitindo de forma eficiente a recuperação da informação, oferecendo acesso ilimitado, diminuindo custos com impressão, facilitando e modificando o fluxo contínuo da informação científica (BOMFÁ; CASTRO, 2004). De acordo com Gonçalves e Barbosa (2015, p. 703), “ a literatura contrapõe o universo do impresso ao universo do digital e apresenta o primeiro como sendo linear e o segundo como não-linear”. Desta forma, o texto disponibilizado pelas redes de computadores, composto por conexões e que pode ser acessado por máquina e por usuários em qualquer lugar e simultaneamente, pode ser definido como hipertexto. Os hipertextos teriam uma estruturação em rede, lidos em objetos não-lineares e possibilitariam práticas de leitura também não-lineares. Passarelli (2008, p.7) destaca que:

O hipertexto institui a possibilidade da narrativa não-linear com os links de trechos de um documento para outros documentos (que podem ser infinitos), desconstruindo a narrativa linear instituída em 500 anos de comunicação por meio da palavra escrita impressa em papel. Desta forma o hipertexto aliado às potencialidades da conexão em rede na Web imprime à informação na sociedade contemporânea em rede características como: instantaneidade, transitoriedade, interoperabilidade e interatividade. (PASSARELLI, 2008, p.7).

Os modelos de publicação tradicionalmente aceitos pela comunidade científica têm se mostrado defasados para as novas formas de se comunicar ciência. Neste cenário, a Ciência Aberta vem possibilitando novos métodos para a divulgação, compartilhamento e reuso, proporcionando um aumento significativo da acessibilidade e da visibilidade. Nesse seguimento, novas práticas para publicação estão sendo desenvolvidas devido ao uso intensivo de dados como, por exemplo, as publicações ampliadas, que possui como características a ampliação do escopo de uma publicação tradicional, ligando-a aos dados provenientes da própria pesquisa, como também outros documentos e objetos digitais relacionados. Conforme Araya (2014, p. 120), “[...] a função de ligação/enlace dos elementos informacionais com a narrativa que trata sobre os resultados da pesquisa é a principal característica da publicação ampliada”. Esta prática possibilita o compartilhamento de dados, aumenta a visibilidade da pesquisa e contribui para o movimento mundial de acesso livre à produção de pesquisa na internet. Sayão e Sales (2016, p.69) afirmam que:

Uma publicação ampliada é uma instância de um objeto digital complexo que se caracteriza pela vinculação de artigos aos dados que subsidiaram a sua criação, bem como à outras informações que auxiliem na contextualização da pesquisa, por exemplo, informações sobre os autores, sobre os equipamentos utilizados, projetos, etc. Essas vinculações podem ter valores

semânticos e estarem associados à diversas ontologias.

A publicação ampliada pode ser considerada uma nova forma de publicação, onde a forma tradicional (um relatório, um artigo, um livro) é enriquecida com informações adicionais, dependente das possibilidades de ligações-conexões da Web (MUCHERONI; SILVA; PALETTA, 2015). No modelo de publicação ampliada, uma publicação científica pode incluir: a) os próprios dados da pesquisa para evidenciar o estudo; b) materiais adicionais usados para ilustrar ou esclarecer a pesquisa; c) dados posteriores à publicação, tais como comentários, classificações etc. (ARAYA, 2014). Já segundo Verhaar (2008, p. 11, *tradução nossa*) os dados de pesquisa:

[...] estão cada vez mais sendo armazenados em repositórios de dados confiáveis. O objetivo da curadoria de dados é garantir que os materiais acadêmicos e científicos possam ser preservados e reutilizados. Entretanto, na atual infraestrutura de comunicação científica, esses conjuntos de dados não estão conectados às publicações onde são discutidos. As publicações ampliadas são criadas com o objetivo de ligar os conteúdos dos repositórios institucionais aos conteúdos dos repositórios de dados.

Desta maneira, o potencial dos repositórios de dados de pesquisa quando interligados a publicações científicas é elevado, pois esses dados expostos dentro de um contexto mais amplo e permitindo aos usuários ter uma visão, que se perde quando os conjuntos de dados são vistos isoladamente (CLARIVATE ANALYTICS, 2017).

Uma publicação tradicional oferece uma interpretação dos dados coletados; e o dado em si traz a sustentação às argumentações e análises. Desta forma, conclui-se que os dados abertos de pesquisa ligados às publicações tradicionais de maneira integrada garantem a identificação de sua autoria como também agrega valor a uma nova pesquisa.

No campo da ciência, não basta somente gerar e disponibilizar uma quantidade maior desses elementos. Para poder atribuí- lhes valor científico, eles devem ser comunicados formalmente como, por exemplo, em periódicos

científicos. Elementos informacionais desvinculados do processo de pesquisa que os originou perdem muito do seu valor informacional (ARAYA, 2014). Para que haja essa integração, o conjunto de dados de pesquisa deve carregar um identificador associado a ele pelo sistema de informação que compõe a agregação/composição, ou seja, um identificador persistente como, por exemplo, o *Digital Object Identifier System* (DOI).

Para configurar-se em um modelo de publicações ampliadas, é importante seguir alguns requisitos e recomendações para desenvolver uma infraestrutura compatível com os padrões complexos associados às publicações ampliadas. Verhaar (2009) descreve alguns desses requisitos: deve ser possível gerar um identificador persistente para cada objeto; os objetos de uma publicação ampliada devem estar disponíveis com recursos da web para que possam ser consultados através de URIs (*Uniform Resource Identifier*); deve ser possível vincular objetos digitais à publicação; deve-se ter o registro e controle de versões da publicação e das partes constituintes; deve ser possível registrar propriedades básicas da publicação e dos recursos que são adicionados a ele (títulos, breve descrição, data da última modificação, data de criação, etc.); registrar a autoria da publicação e dos objetos digitais: as pesquisas científicas são cada vez mais colaborativas, por este motivo deve-se registrar as contribuições individuais; garantir a preservação a longo prazo das publicações ampliadas; deve ser possível registrar as relações entre os recursos web que fazem parte da publicação ampliada (os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão); as instituições que oferecem acesso às publicações ampliadas devem manter os metadados dos objetos em versão legível por máquinas para que sejam descobertos; as instituições que fornecem acesso às publicações ampliadas devem certificar-se de que estejam disponíveis como documentos baseados no modelo OAI-ORE (*Open Archives Initiative - Object Reuse and Exchange*).

Descrevemos aqui alguns dos requisitos que auxiliam no desenvolvimento de uma infraestrutura compatível com as publicações ampliadas. Contudo considera-se necessário um aprofundamento a partir de outros

estudos sobre os requisitos necessários para a composição de um modelo. Nota-se a necessidade de esquemas mais ricos de metadados e uma maior complexidade das infraestruturas que estão permeando os dados, diferente do gerenciamento das publicações digitais convencionais. As publicações convencionais apresentam características como, por exemplo, documentos em PDF; textos em HTML; textos com links onde os dados não são curados; textos em desacordo com normas e padrões relacionados a Web Semântica e textos não integrados com outras informações de outros repositórios. (SALES, 2014), características estas que diferem do modelo de publicações ampliadas.

4 UM PANORAMA DE INTEGRAÇÃO ENTRE REPOSITÓRIOS DE PESQUISA E PUBLICAÇÕES CIENTÍFICAS BASEADO NO MODELO DE PUBLICAÇÕES AMPLIADAS

Na primeira fase de análise, foram coletadas literaturas específicas da área por meio da revisão sistemática. A partir da literatura, foram extraídos os requisitos indicados e sugeridos para um modelo de publicação ampliada. Na segunda fase de análise, foram selecionados os repositórios de dados de pesquisa e averiguadas as informações disponibilizadas por meio dos filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada. Na terceira fase de análise, foram armazenados os *screenshots* das interfaces dos repositórios coletados na segunda fase, que utilizam o modelo de publicações ampliadas.

A revisão sistemática da literatura, realizada na primeira fase da análise, em bases de dados eletrônicas utilizando um *string* de busca, recuperou um total de 11 itens a serem analisados, sendo que somente cinco documentos abordavam assuntos relacionados à infraestrutura de uma publicação ampliada e seus principais requisitos. Mesmo com um grande empenho para desenvolver filtros de pesquisa que recuperassem estudos relevantes com maior precisão, observou-se um número

baixo de documentos referentes ao assunto *enhanced publication* e que tratassem sobre seus principais requisitos. Desta maneira, após a seleção dos estudos, os dados foram extraídos e sintetizados para que fosse possível estabelecer as categorias para análise.

Além dos requisitos coletados durante a revisão sistemática, também foram incluídos nas categorias os requisitos citados por Verhaar (2009), bem como os estudos referenciados nas obras recuperadas durante a revisão sistemática, sendo assim, as categorias para análise foram estabelecidas.

Na segunda fase da análise foram selecionados os repositórios de dados de pesquisa através do diretório *re3data.org* e averiguadas as informações disponibilizadas por meio dos filtros que demonstrassem a utilização de ligações baseadas no modelo de publicação ampliada. Foram descritos os tipos de conteúdos, quantidades de repositórios por países, identificadores digitais, tipos de padrões de programação, restrições de acesso aos dados, formas de upload, tipos de instituições vinculadas aos repositórios, padrões de metadados, idiomas, softwares utilizados e a quantidade destes repositórios que utilizam publicações ampliadas, totalizando 654 repositórios.

Na terceira fase da análise foi realizada a inspeção das interfaces e identificados os elementos que cada repositório contempla a partir dos requisitos estabelecidos por meio da revisão sistemática. Posteriormente iniciaram-se as análises, observações e comparações dos elementos utilizados nos repositórios, destacando-se os principais recursos. Entre os 654 repositórios definidos pelo diretório, com publicações ampliadas, somente 136 foram inseridos no estudo e 518 repositórios foram excluídos por motivos diversos como, por exemplo, publicações não relacionadas aos conjuntos de dados de pesquisa.

Entre os 136 repositórios selecionados foram evidenciados os recursos disponíveis como as áreas e formas mais abrangentes, destacando-se a Biologia, Ciências Naturais e Medicina. Observou-se que as formas de dados mais disponibilizadas entre os repositórios de dados de pesquisa com publicações ampliadas são de dados estatísticos, em texto e em imagens.

Em relação aos identificadores digitais de autoria, foi concluído que um grande índice

de repositórios de dados de pesquisa com publicações ampliadas ainda não utiliza um identificador digital persistente para autor. Contudo, entre os padrões de programação destacam-se os padrões REST e OAI-PMH.

Referente às restrições de acesso aos dados constatou-se que um percentual maior dos repositórios selecionados possibilita o acesso por meio de registro ou através de alguma associação institucional e um pequeno percentual através de depósito de dados e de taxa requisitada. Sobre as instituições vinculadas aos repositórios, observou-se que grande parte estão vinculados a instituições sem fins lucrativos.

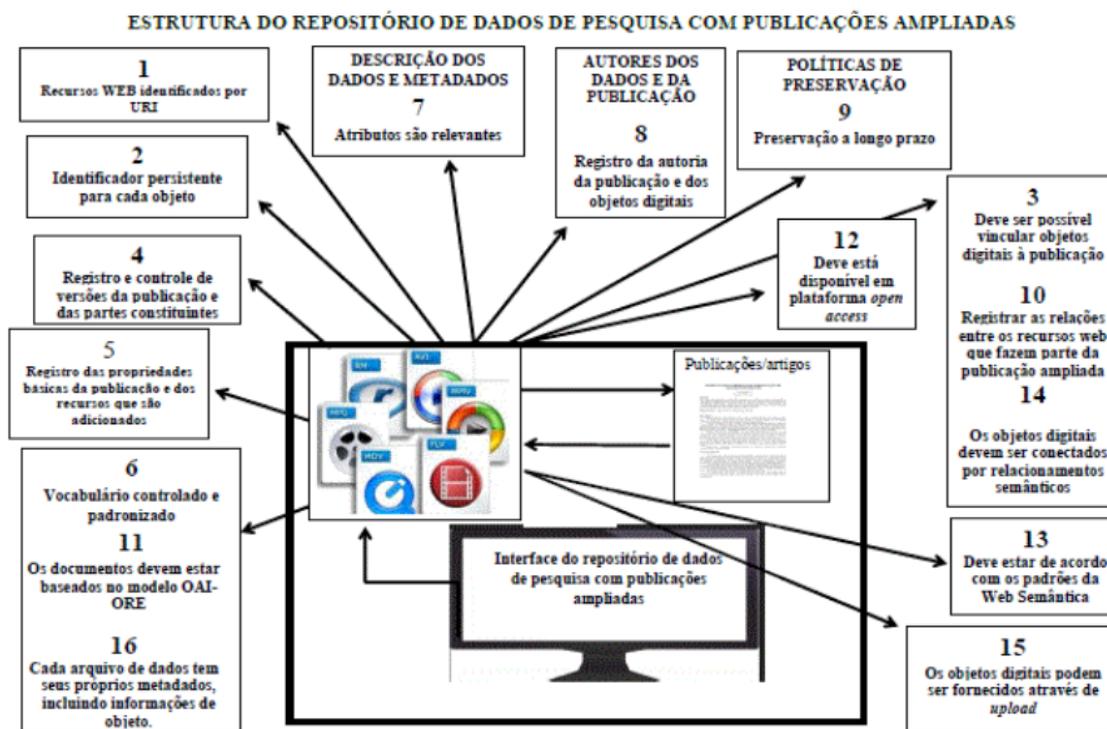
Para os padrões de metadados, foi identificado uma taxa de 21% de repositórios que utilizam DDI; 16% utilizam Dublin Core e DataCite; 4% utilizam ISA e apenas 3% utilizam RDF e *Repository Developed Metadata Schemas*. Conforme informações do diretório, a maior parte dos repositórios utilizam o identificador digital de objetos digitais DOI ou não utilizam nenhum tipo de identificador.

O principal idioma destacado entre os 136 repositórios é o inglês, com 98% e o restante dos idiomas como, por exemplo, o Alemão, Sueco, Francês, Espanhol, Polonês e Finlandês aparecem com um percentual de 1% dos repositórios. Em relação aos softwares mais utilizados; 12% dos repositórios utilizam o Dataverse e 3% utilizam o DSpace e EPrints. Os tipos de repositórios identificados neste estudo foram os temáticos, com 65% e os institucionais com 30%. Entre os países mais destacados, 47% dos repositórios são provenientes dos Estados Unidos; 10% da Alemanha e Reino Unido; 6% da União Europeia; 5% do Canadá e Austrália. Entre outros países, há um percentual de 32% de repositórios.

Após serem evidenciados os recursos disponíveis entre os 136 repositórios, foram destacados os 16 requisitos de publicações ampliadas definidos nesta pesquisa. As interfaces dos repositórios foram inspecionadas para a verificação da ligação de seus dados e se contemplavam os 16 requisitos.

Cruzando as informações existentes na literatura com os dados coletados nos repositórios, foi proposto um panorama de integração para repositórios de dados de pesquisa e as publicações científicas relacionadas, baseando-se na publicação ampliada, conforme demonstra a figura 2:

Figura 2: Estrutura da publicação ampliada para integração de dados de pesquisa



Fonte: Elaborado pelos autores

Uma prática indicada na literatura consiste em associar identificadores globais únicos e persistentes com cada um dos recursos, conforme destacado no requisito número dois. Na inspeção verificou-se que entre os 136 repositórios inseridos no estudo, 81 destes apresentam identificadores persistentes para objetos digitais e o mais destacado entre eles foi o DOI.

O requisito três descreve a possibilidade de vincular objetos digitais à publicação, ou seja, as publicações ampliadas são consideradas objetos altamente complexos, que agrupam os textos de artigos de revistas, imagens, arquivos de vídeos, etc. Durante a inspeção foi possível identificar que 120 repositórios estão de acordo com o requisito número dois e vinculam objetos digitais às publicações. No restante dos repositórios observou-se que apenas existem referências descritas, mas sem nenhum link de ligação, desta maneira entende-se que não havia vínculos de forma clara.

O requisito número quatro trata sobre o registro e controle de versões da

publicação e das partes constituintes. Dentre os 136 repositórios selecionados para esta pesquisa, 77 apresentaram o requisito quatro. Recomenda-se registrar as propriedades básicas da publicação e dos recursos que são adicionados a ele como, por exemplo, títulos, breve descrição, data da última modificação e data de criação. Estas características estão relacionadas ao requisito número cinco. A análise resultou em 108 repositórios que contemplam este requisito. Foram constatados 60 repositórios que utilizam um vocabulário para registrar a relação entre os objetos digitais, conforme o requisito seis. Conforme citado neste estudo, atributos são relevantes, isto é, recomenda-se que cada componente seja digitado semanticamente para deixar claro que tipo de recurso está sendo referido. Esta característica está contemplada no requisito número sete e observou-se que 90 repositórios possuíam uma breve descrição de cada componente inserido ao conjunto de dados, o que possibilitou a identificação sobre

que tipo de arquivo estava disponível antes de fazer o download. Entre os 46 repositórios que não contemplavam o requisito número sete, foi verificado que grande parte não utilizava metadados como também não havia descrições de cada objeto digital, fornecendo somente o link de acesso para download.

De acordo com as análises, 105 repositórios possuíam o registro da autoria da publicação e dos objetos digitais e estavam de acordo com o requisito número oito. Constatou-se que em alguns repositórios havia os registros das contribuições individuais e a distinção clara entre o autor da publicação científica e os autores de suas partes componentes.

Um total de 126 repositórios não contemplava o requisito número nove, que trata sobre políticas de preservação. O motivo se deu pela falta de identificação em suas interfaces de informações relacionadas à preservação. Acredita-se que as instituições possam ter documentos internos que relatem suas políticas de preservação, no entanto nesta pesquisa não foram reconhecidos. Os 10 repositórios restantes apresentaram algum tipo de documentação ou instruções referentes à preservação.

O requisito 10 refere-se ao registro das relações entre os recursos web que fazem parte da publicação ampliada, isto é, os componentes precisam ser descritos e classificados utilizando-se um vocabulário padrão. Ao longo das análises dos 136 repositórios, foi possível observar que 103 deles possuíam, visivelmente, as relações entre os recursos web que faziam parte da publicação ampliada. O requisito 11 recomenda que os documentos estejam baseados no modelo OAI-ORE. A partir desta característica, verificou-se que entre os repositórios selecionados, 78 estavam de acordo com este requisito.

Diferentes componentes que fazem parte de uma publicação ampliada podem ter distintos tipos de acesso e se tornar inacessíveis, por este motivo, é importante que os dados estejam disponíveis em plataformas *open access*, conforme recomendado no requisito número 12. Percebeu-se que praticamente, a totalidade dos repositórios está de acordo com o requisito.

Estar de acordo com os padrões da Web Semântica é o que recomenda o requisito número 13. Conforme já mencionado, a web semântica tem como objetivo organizar a informação de maneira legível por máquinas através de

padrões de formatação de dados. A partir dessa recomendação, observou-se na inspeção que 55 repositórios apresentavam as características descritas neste requisito.

O requisito 14 recomenda que os objetos digitais estejam conectados por relacionamentos semânticos. Foi percebido nesta pesquisa que este requisito assemelha-se aos requisitos número três e 10 e estabelece que, para ser considerada uma publicação ampliada, os objetos digitais devem estar relacionados à publicação e vice-versa. Estas relações devem estar visíveis através de links e de descrições. Portanto, verificou-se que 121 repositórios estavam de acordo com o requisito 14.

Os objetos digitais podem ser fornecidos através de *upload*, é o que estipula o requisito número 15, ou seja, a ação de enviar dados aos repositórios de dados de pesquisa. Alguns repositórios possuem visivelmente em suas interfaces um link para envio após o cadastramento do usuário no sistema e neste estudo destacou-se que 93 repositórios não possuíam em suas interfaces este item. Apenas forneciam um contato para solicitação de envio.

Referente ao requisito 16, cada arquivo de dados tem seus próprios metadados, incluindo informações de objeto. Entre os 136 repositórios, 72 apresentaram metadados e informações do objeto, considerando os repositórios que possuíam a descrição dos metadados de forma visível. Notou-se que este requisito assemelha-se ao requisito número 6 e ao requisito número 11.

Durante a inspeção das interfaces, foram identificados os elementos que cada repositório contempla para posteriormente iniciar as análises, observações e comparações dos elementos utilizados nos repositórios com os requisitos estabelecidos por meio da revisão sistemática. As observações foram registradas no *checklist* de validação de requisitos.

A partir dos registros das observações no *checklist* de validação, foram elaborados quadros comparativos para verificação dos requisitos essenciais e comuns utilizados entre os repositórios. Com o auxílio dos instrumentos utilizados para a coleta, foi possível dar início as análises dos dados. As informações coletadas, que foram registradas, foram analisadas e comparadas para que fosse possível evidenciar

os elementos essenciais para o desenvolvimento de um modelo padrão baseado na publicação ampliada.

Pode-se compreender, portanto, que a integração de dados de pesquisas às publicações científicas utilizando-se os requisitos associados às publicações ampliadas consequentemente melhora a recuperação da informação de modo que as informações relacionadas se tornem de fácil acesso, além disto, permite uma melhor interpretação dos dados vinculados podendo ser reutilizados em outros estudos. Observou-se, com o apoio da revisão sistemática, que os estudos referentes à publicação ampliada ainda carecem de maior detalhamento e explanação de ações, o que confirma o estágio inicial que se encontra o tema. A partir de definições abordadas na literatura e após as análises dos repositórios concluiu-se que a utilização de requisitos referentes à web semântica permite realizar a busca da informação de forma mais ampla devido às várias relações, isto é, as informações não estavam disponíveis de forma isolada. Requisitos que abordaram os elementos de relacionamentos proporcionaram a integração; e os requisitos equivalentes aos registros de autoria, descrições e controle de versões ofereceram qualidade destacando os direitos autorais, ou seja, aumentaram a visibilidade tanto de publicações, quanto de autorias.

5 CONSIDERAÇÕES FINAIS

O panorama proposto mostrou uma das formas possíveis de apresentação de uma interface de repositórios de dados de pesquisa caracterizada com publicação ampliada a partir de requisitos estabelecidos por meio da literatura recuperada na revisão sistemática.

Definimos manter o princípio de que uma publicação tradicional deve estar relacionada aos objetos digitais para ser considerada uma publicação ampliada. Percebe-se na literatura que há uma tendência para seguir esta definição mais ampla, no entanto existem diversas características a serem consideradas para que, de fato, possa estar de acordo com os critérios associados a este tipo de publicação.

A pesquisa fundamentou-se em investigar quais elementos são importantes nas interfaces de repositórios de dados de pesquisa para que se tenha, de fato, um modelo de publicação ampliada. Constatou-se nas análises que praticamente a totalidade dos repositórios possuem o requisito um, ou seja, tinham os recursos web identificados por URI. Conforme consta na literatura é importante constatar se os manuscritos e dados de pesquisa que são depositados em repositórios fazem parte da arquitetura web, o que implica que todos esses objetos deveriam estar disponíveis como recursos da web que podem ser referenciados através de URI. Para que fosse possível exemplificar cada um dos requisitos encontrado na análise, utilizou-se alguns *screenshots* para as demonstrações.

Foi possível perceber que os repositórios que contemplavam grande parte dos requisitos estabelecidos nesta pesquisa tinham usabilidade, tornando as interfaces agradáveis, eficazes e eficientes durante a realização das buscas. Ou seja, as buscas executadas foram de forma completa e com um menor tempo gasto para a recuperação, diferente das interfaces dos repositórios que não possuíam diversos requisitos.

Assim sendo, a partir do desenvolvimento das três etapas de análise deste estudo, pode-se ter uma visão geral para que fosse possível desenvolver um panorama de integração entre repositórios de dados de pesquisa e publicações científicas baseado no modelo de publicações ampliadas. Espera-se que a presente pesquisa possa contribuir para a ampliação do conhecimento referente ao tema e servir de referência para o desenvolvimento de repositórios de dados de pesquisas que possuem o intuito de implementar este tipo de publicação.

Novas investigações poderão contribuir para o desenvolvimento de modelos e padrões associados às publicações ampliadas, desta maneira, outras propostas de infraestruturas e modelos podem ser sugeridas, assim como avaliações de interfaces de repositórios que já contemplam os requisitos associados às publicações ampliadas.

Artigo recebido em 01/03/2018 e aceito para publicação em 03/05/2018

ENHANCED PUBLICATIONS: aspects of integrating search data

ABSTRACT This article presents the results of a research that had as main objective propose a panorama of integration between repositories of research data and scientific publications, based on the model of enhanced publications. The analysis methodology of this study was composed of three phases, starting with the literature review of the area, then identifying repository data from the re3data.org directory, using links based on the extended publication model and, lastly, the inspection of the interfaces of selected repositories. It was verified a growing need to make use of an infrastructure compatible with the enhanced publications, making it important the use of requirements for the interfaces of the research data repositories. The cross-checking of information carried out in this study allowed a better understanding of how relevant the use of each of the studied requirements was, and it also allowed to understand the consequences when these requirements are not included in the repositories interfaces. From the research and the deeper study on the subject, it was possible to broaden the knowledge and detect the appropriate elements to propose a panorama of integration between a repository of research data and scientific publications based on an enhanced publications model, as well as highlighting the advantages of using this model.

Keywords: Enhanced publications. Research data repositories. Requirements for enhanced publications.

REFERÊNCIAS

- ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Orgs.). **Ciência aberta: questões abertas**. Brasília: IBICT, 2015.
- ARAYA, Elizabeth Roxana Mass. **Comunicação científica: agregação, compartilhamento e reuso de elementos informacionais**. 2014. Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação. Faculdade de Filosofia e Ciências. Universidade Estadual Paulista. Marília, 2014. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/121981/000812425.pdf?sequence=1&isAllowed=y>>. Acesso em: 2 maio, 2017.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, [19--].
- BERLIN **Declaration on open access to knowledge in the sciences and humanities**. Berlin, 2003. Disponível em: <<https://openaccess.mpg.de/Berlin-Declaration>>. Acesso em: 11 mar. 2017.
- BOMFÁ, Cláudia Regina Ziliotto; CASTRO, João Ernesto E. Desenvolvimento de revistas científicas em mídia digital: o caso da Revista Produção on-line. **Ciência da Informação**, v. 33, n. 2, dez. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a04v33n2>>. Acesso em: 26 out. 2017.
- CASTEDO, Raquel da Silva. Periódicos científicos on-line: novas interfaces, novos usos, novas práticas. In: XXX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO – Santos, 2007. **Anais...** Santos, 2007. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2007/resumos/R1485-1.pdf>>. Acesso em: 27 out. 2017.
- GONÇALVES, Márcio Souza; BARBOSA, Rafael de Oliveira. Comunicação, linearidade e não linearidade: costurando conceitos e práticas. **Intexto**, n. 34, p. 694-712, set./dez. 2015. Disponível em: <<http://seer.ufrgs.br/index.php/intexto/article/viewFile/58500/35515>>. Acesso em: 26 out. 2017.
- IBICT. **IBICT Dataverse**. 2017. Disponível em: <<https://repositoriopesquisas.ibict.br/dataverse/ibict>>. Acesso em: 31 ago. 2017.

- KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. Joint Technical Report Software Engineering Group, Department of Computer Science Keele University, United King and Empirical Software Engineering, 2004. Disponível em: <<http://www.ifs.tuwien.ac.at/~weipl/systemicReviewsSoftwareEngineering.pdf>>. Acesso em: 08 set. 2017.
- NASSI-CALÒ, Lilian. Teses e dissertações: prós e contras dos formatos tradicional e alternativo. **SciELO em perspectiva**, 2016. Disponível em: <<http://blog.scielo.org/blog/2016/08/24/teses-e-dissertacoes-pros-e-contras-dos-formatos-tradicional-e-alternativo/#.WpBkDvnwZdh>>. Acesso em: 27 out. 2017.
- MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.
- MUCHERONI, Marcos Luiz; SILVA, Fernando José Modesto da.; PALETTA, Carlos Francisco. Entre a publicação ampliada e a multimodalidade. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 16., 2015, João Pessoa. Anais... João Pessoa, 2015.. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000017566/524e08c83118ed11f73/>>.
- PASSARELLI, Brasilina. Do Mundaneum à WEB Semântica: discussão sobre a revolução nos conceitos de autor e autoridade das fontes de informação. **DataGramZero**, v.9 n.5 out. 2008. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/download/7638>>. Acesso em: 24 out. 2017.
- RNP- Rede Nacional de Ensino e Pesquisa. Disponível em: <<https://www.rnp.br/>>. Acesso em: 09 set. 2017.
- SALES, Luana Farias; SAYÃO, Luís Fernando; SOUZA, Rosali Fernandez. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-science. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 2013, Florianópolis. **Anais ...** Florianópolis, 2013. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000013705/54d555cb665c858479ea1dfd3f417ff/>>. Acesso em: 23 jul. 2016.
- SANTOS, Paola. Paul Otlet: um pioneiro da organização das redes mundiais de tratamento e difusão da informação registrada. **Ciência da Informação**, v. 36, n. 2, p. 54-63, maio/ago. 2007. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1176/1339>>. Acesso em: 24 out. 2017.
- SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS |** 2014, jun., 8(2) - p.76-92, 2014. Disponível em: <<http://www.reciis.icict.fiocruz.br/>>. Acesso em: 02 ago. 2016.
- _____. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**. v.5, n.2, p. 67-71, 2016.
- SOUZA, Ricardo Augusto Teixeira de. **Predição de tags usando linked data**: um estudo de caso no banco de dados Arquigrafia. 2013. Dissertação (Mestrado) - Programa de PósGraduação em Ciência da Informação. Faculdade de Filosofia e Ciências. Universidade Estadual Paulista. Marília, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45134/tde-04062014-230955/pt-br.php>>. Acesso em: 11 ago. 2016.
- VERHAAR, Peter. Object models and functionalities VERNOOY-GERRITSEN, Marjan et al. In: **Enhanced Publications: linking publications and research data in digital repositories**. SURF Foundation: Amsterdam University Press, Amsterdam, 2009.

