

Repositórios institucionais de acesso aberto: adequação às novas métricas da web

Institutional open access repositories: adapting to new web metrics

Repositorios institucionales de acceso abierto: adecuación a las nuevas métricas de la web

Ana Maria Ferreira de Carvalho | anacarvalho@facc.ufrj.br

Universidade Federal do Rio de Janeiro, Escola de Comunicação, Programa de Pós-graduação em Ciência da Informação. Rio de Janeiro, Brasil.

Fabio Castro Gouveia | fabio.gouveia@fiocruz.br

Fundação Oswaldo Cruz, Casa de Oswaldo Cruz, Museu da Vida. Rio de Janeiro, Brasil.

Resumo

Repositórios institucionais de acesso aberto são elementos essenciais no sistema de comunicação científica da atualidade. Observa-se crescimento significativo de seu uso por instituições públicas de ensino superior brasileiras, nos últimos quinze anos. Cresce o uso do ambiente da web para produção, armazenamento, disseminação e acesso à informação científica e tecnológica, impactando toda a estrutura da comunicação científica. Este novo formato carece de novos indicadores para os estudos métricos da informação científica e tecnológica, como os indicadores webométricos e altmétricos. Por meio de pesquisa exploratória e descritiva, apresenta como estas instituições estão adequando seus repositórios a este novo cenário. Foram identificados que 70% dos repositórios disponibilizam estatísticas de acesso e *download* de itens, 22% disponibilizam dados altmétricos, 83% utilizam identificadores persistentes de objetos digitais. Mostrou-se que as instituições públicas de ensino superior brasileiras estão adequando seus repositórios ao novo cenário dos estudos métricos da informação, apresentando, a altmetria, resultado ainda incipiente.

Palavras-chave: Repositórios Institucionais; Acesso Aberto; Métricas da Web; Comunicação Científica; Indicadores.

Abstract

Institutional open access repositories are essential in today's scholarly communication system. It's observed significant growth of its use by public institutions of higher education in Brazil in the last fifteen years. The use of the web environment for production, storage, dissemination and access to scientific and technological information grows, impacting the entire structure of scientific communication. This new format needs new indicators for the metric studies of scientific and technological information, as webometric and altmetrics indicators. Through exploratory and descriptive research, it shows how these institutions are adapting their repositories to the new scenario. It was identified that 70% of the repositories provide statistical data on access and download of its items, 22% offer altmetric data, 83% use "identifiers resolution system" of digital objects. It shows that public higher education institutions are now adapting their repositories to the new scenario of metric information studies, and altmetrics presents incipient results.

Keywords: Institutional Repositories; Open Access; Web Metrics; Scientific Communication; Indicators.

Resumen

Repositorios institucionales de acceso abierto a los elementos básicos no sistema de comunicación científica de la actualidad. Se observa crecimiento significativo de su uso por las instituciones públicas de superior brasileiras, nos últimos quince años. Creación de un sitio web para la producción, almacenamiento, difusión y acceso a la información científica y tecnológica, impacto en toda la estructura de la comunicación científica. Este nuevo formato de nuevos indicadores para los estudios técnicos de la información científica y tecnológica, como los indicadores webométricos y altmétricos. Por medio de la investigación exploratoria y descriptiva, se presentan como estas entidades que están adecuando sus repositorios a este nuevo escenario. Hemos identificado que el 70% de los depósitos disponibles y las estadísticas de acceso a los ítems, 22% disponen datos altmétricos, 83% utilizan identificadores persistentes de objetos digitales. Se muestra que como instituciones públicas de enseñanza superior brasileiras están adecuando sus repositorios al nuevo escenario de los estudios métricos de la información, presentando, a altmetria, el resultado aún incipiente.

Palabras clave: Repositorios Institucionales; Acceso Abierto; Métricas da Web; Comunicación Científica; Identificadores.

Introdução

Os Repositórios Institucionais (RI) de acesso aberto ganham destaque no sistema de comunicação científica da atualidade. Impulsionado pelo desenvolvimento das Tecnologias de Informação e Comunicação (TIC) e pela crise que se instalava no mundo, na última década do século XX, em relação ao acesso a informações científicas e tecnológicas, os RI de acesso aberto vêm ganhando força nos últimos quinze anos. Sua implantação, em diversas instituições de ensino e pesquisa, tem como principal objetivo o armazenamento, a disseminação, a preservação e a promoção de maior visibilidade à produção científica de uma instituição.

Historicamente, os primeiros repositórios utilizados pela comunidade científica foram os repositórios temáticos, seguidos pelos repositórios de teses e dissertações e mais especificamente, os repositórios institucionais, que se tornaram uma via alternativa para a comunicação científica.

O Movimento de Acesso Aberto (*Open Access Movement*), que impulsionou a implantação de repositórios institucionais, representou uma reação ao momento histórico conhecido como a crise dos periódicos científicos. Marcado pelos altos custos das assinaturas desses periódicos e bases de dados, inviabilizava, principalmente nos países menos desenvolvidos, o acesso à produção científica, muitas vezes financiada por recursos públicos.

O advento da Internet, e principalmente da web, permitiram que os processos de comunicação científica se dessem de forma mais rápida e colaborativa. Neste cenário, os repositórios institucionais de acesso aberto têm papel estratégico como facilitadores de acesso à produção científica, fazendo parte do fluxo da comunicação científica. Observa-se que um dos principais documentos depositados nos RI de acesso aberto são artigos de periódicos revisados por pares. Esta literatura ganha maior visibilidade quando disponível de forma livre e gratuita, podendo obter, desta forma, mais citações dentro de uma comunidade científica.

Para avaliar a qualidade dos processos de produção de conhecimento científico, os estudos em Biblioteconomia e na Ciência da Informação fazem uso de indicadores, que permitem entender com se dá sua evolução e auxiliar na tomada de decisões que envolvem o desenvolvimento da pesquisa em uma área do conhecimento. A bibliometria e a cientometria são as técnicas mais utilizadas nesses estudos, mas novos campos de estudos métricos surgem, considerando o contexto da *Web*. Para observarmos e mensurarmos os processos e os fenômenos relacionados à comunicação científica que se originam neste contexto, faz-se necessário o uso de métricas e de indicadores que se adéquem a este novo cenário, dando subsídios às análises.

A webometria, a webmetria e a altmetria são exemplos de métricas criadas com este objetivo, tendo o *link*, as estatísticas de acesso, de visitas e de visualização da informação, assim como as menções em redes sociais, como indicadores, estas métricas buscam avaliar, a visibilidade, a popularidade e o prestígio dos autores e de suas produções, entre outras possibilidades.

O estudo e o uso das métricas da *Web* para a análise do fluxo de informação e das relações que se estabelecem na rede vêm crescendo sensivelmente nas comunidades científicas. Há, atualmente, várias ferramentas que capturam e consolidam informações que permitem a realização de estudos dessas relações. Observa-se, também, que vários *sites* de RI estão se adequando a este novo cenário, disponibilizando dados e informações que permitam a realização de estudos webmétricos e altmétricos sobre seus acervos e seus pesquisadores.

O presente estudo tem como objetivo apresentar os principais conceitos e relações entre as métricas da informação desde as mais tradicionais, até as denominadas alternativas, direcionadas para o ambiente da *Web*. Também observar, por meio de levantamento de dados nos *sites* dos RI de acesso aberto de IPES brasileiras, como estas instituições estão atentas aos novos indicadores de avaliação da produção científica em ambiente web e como estão se adequando a este novo formato, disponibilizando nos *sites* de seus RI elementos que são utilizados para estudos webométricos e altmétricos, que permitam avaliar o interesse por

seu conteúdo, sua visibilidade, prestígio dos autores, entre outras informações relevantes sobre a produção científica da instituição.

Para tanto, nas seções que se seguem serão apresentados alguns conceitos sobre repositórios institucionais e as questões que envolvem o Movimento de Acesso Aberto, os conceitos das principais métricas tradicionais e alternativas da Ciência da Informação, mostrando suas relações, para, em seguida, apresentar os dados que visam mostrar se os RI das IPES brasileiras estão se adequando às novas métricas, ou métricas alternativas, para o ambiente da Web.

Repositório institucional de acesso aberto

Os Repositórios Institucionais (RI) de acesso aberto são uma das principais estratégias do Movimento de Acesso Aberto, denominada Via Verde. O desenvolvimento e a implantação de repositórios institucionais buscam ampliar os esforços para eliminar barreiras que comprometam o acesso à produção científica, promovendo maior visibilidade das atividades de pesquisa, contribuindo, assim, para o desenvolvimento científico e tecnológico de uma comunidade ou de um país. Neste sentido, os RI, além de seu papel político, como ressaltado por Sayão e colaboradores¹, são fundamentais para a gestão da informação científica e tecnológica, na sua manutenção e preservação, como será apresentado a seguir, considerando os conceitos expressos por autores estudiosos do tema.

Crow², em trabalho de destaque sobre o tema, introduz o conceito de repositório institucional e apresenta o que considera serem os elementos essenciais para a implantação de tal recurso tecnológico. O autor esclarece que, “em termos gerais, um repositório digital institucional refere-se a qualquer coleção de material digital hospedado, de propriedade, ou controlado, ou divulgado por uma faculdade ou universidade, independentemente do propósito ou proveniência”¹ (tradução nossa)². Destaca o repositório institucional como a coleção digital para captura e preservação da produção intelectual de uma única comunidade científica ou de conjuntos de comunidades. Considera que o uso deste recurso, entre outras vantagens, expande o acesso aos resultados de pesquisa, oferece maior controle sobre os processos de produção científica de sua instituição, reduz o poder de editoras privadas sobre o controle da produção científica, aumentando a relevância do papel das instituições acadêmicas e das bibliotecas nos processos de produção de suas comunidades científicas. O autor complementa afirmando que a inserção desse recurso tecnológico nos processos de produção e disseminação dos resultados de pesquisa causa relevante mudança nos processos de comunicação científica e são potenciais indicadores de qualidade de uma instituição acadêmica; dando maior visibilidade as suas produções, incrementando seu prestígio e seu valor público. Para Crow² é essencial que o conteúdo dos repositórios institucionais seja exclusivamente acadêmico, definido por sua comunidade, aliado ao fato de que deve ser cumulativo e perpétuo, aberto e interoperável.

Enfatizando o papel relevante do repositório institucional para as universidades, Lynch³, os define com um conjunto de serviços que estas oferecem a seus membros, para o gerenciamento e disseminação do conteúdo digital produzido pela instituição. O autor também chama a atenção para as questões relacionadas à administração, à criação de políticas, à preservação, ao acesso a esses materiais, destacando o papel do bibliotecário, dos profissionais de tecnologia da informação e dos arquivistas como colaboradores para a manutenção desses recursos tecnológicos e seus serviços.

O autor destaca, ainda, o papel estratégico dos repositórios institucionais para as universidades e instituições de pesquisa que, sendo mais um instrumento dentro do sistema de comunicação científica, proporciona novas possibilidades dentro de seus processos, inovando e encorajando a exploração e a adoção de novas formas de comunicar, dando maior visibilidade aos vários resultados obtidos pela instituição, por

¹ *a digital institutional repository could be any collection of digital material hosted, owned or controlled, or disseminated by a college or university, irrespective of purpose or provenance.*

meio das pesquisas, não se limitando somente àqueles que percorrem o caminho tradicional das publicações científicas, bem consolidadas nesse contexto.

Leite⁴ considera que RI representa um dos vários tipos de Repositório Digitais, sendo este um conceito mais amplo que se constitui dos “vários tipos de aplicações de provedores de dados que são destinados ao gerenciamento de informação científica, constituindo-se, necessariamente, em vias alternativas de comunicação científica”⁴. Ainda segundo o autor, RI são voltados à produção intelectual de uma instituição, especialmente universidades e institutos de pesquisa. O autor, por outro lado, amplia a categorização dos repositórios digitais de acordo com o tipo de aplicação provedora de dados, incluindo neste grupo, os repositórios de teses e dissertações. Para Leite⁴, “considera-se que, de maneira geral, os repositórios digitais podem ser de três tipos: 1) *repositórios institucionais*: voltados à produção intelectual de uma instituição, especialmente universidades e institutos de pesquisa. Exemplo: *e-Prints Soton* – repositório de Pesquisa da Universidade de Southampton (<http://eprints.soton.ac.uk/>); 2) *repositórios temáticos* ou *disciplinares*: voltados a comunidades científicas específicas. Tratam, portanto, da produção intelectual de áreas do conhecimento em particular. Exemplo: *E-LIS – EPrints in Library and Information Science* (<http://eprints.rclis.org/>) e *arXiv.org* (<http://arxiv.org/>); 3) *repositórios de teses e dissertações (Electronic Theses and Dissertations – ETDs)*: repositórios que lidam exclusivamente com teses e dissertações. Muitas vezes a coleta das muitas *ETDs* é centralizada por um agregador. Exemplo: *BDTD/UnB – Biblioteca Digital de Teses e Dissertações da Universidade de Brasília* (<http://bdttd.bce.unb.br/tedesimplificado/>) e *BDTD* (<http://bdttd.ibict.br/>)”⁴.

Leite⁴ ressalta a importância e os benefícios dos repositórios institucionais para pesquisadores, para as universidades, para os administradores acadêmicos e para toda a comunidade científica. Seu uso, entre outras possibilidades, permite melhorar a comunicação científica, contribuir para a preservação dos conteúdos digitais, oferecer insumos para avaliação e monitoramento de sua produção, dar maior acesso e maior visibilidade à produção institucional, podendo maximizar seu impacto e contribuir com o prestígio da instituição e de seus pesquisadores.

Estudos como os realizados pelos pesquisadores da Universidade do Minho, em Portugal, Rodrigues, Swan e Baptista⁵ um dos países que se destacam em pesquisa sobre o tema, mostram os inúmeros benefícios que os repositórios institucionais, particularmente os de acesso aberto, proporcionam as suas instituições sendo considerado, atualmente, um serviço de grande sucesso e relevância para sua comunidade acadêmica.

Os dados disponibilizados pelo diretório *OpenDOAR*ⁱⁱ permitem observar um crescimento significativo de seu uso por instituições públicas de ensino superior brasileiras, nos últimos quinze anos, acompanhando o crescimento mundial, com 2.878 RI cadastrados em julho de 2017, sendo destes 74 no Brasil.

O movimento de acesso aberto

O Movimento de Acesso Aberto (*Open Access Movement*) representou uma reação ao momento histórico conhecido como a crise dos periódicos científicos, marcado pelos altos custos das assinaturas desses periódicos e bases de dados, inviabilizando, principalmente nos países menos desenvolvidos, o acesso à produção científica, muitas vezes financiada por recursos públicos. Marcado pela Iniciativa do Acesso Aberto de Budapeste ou *Budapest Open Access Initiative* (BOAI)⁶, o movimento firma a Declaração de Budapeste, onde garante que o acesso livre é viável, tendo em vista algumas iniciativas já implementadas. Se por um lado proporciona aos usuários o “extraordinário poder de encontrar e fazer uso das informações que lhes são relevantes”⁶, por outro, o autor e seu trabalho tem a possibilidade de ter maior visibilidade, maior acesso e, conseqüentemente, maior impacto em sua produção. O documento recomenda duas

ii Diretório de dados sobre repositórios institucionais de acesso aberto no mundo. <http://www.opendoar.org>

estratégias complementares à filosofia e que são os principais mecanismos para a publicação em acesso aberto. A primeira estratégia, que ficou conhecida como Via Verde, trata do autoarquivamento, que se refere às condições necessárias para que o próprio autor possa depositar sua produção científica já publicada ou aceita para publicação, com a autorização dos editores. A segunda estratégia, conhecida como Via Dourada, trata do meio de disseminação da produção científica sem custo para o usuário. Isto é, refere-se aos periódicos de acesso aberto, uma via por meio da qual os pesquisadores e qualquer outro usuário poderão fazer acesso ao conteúdo acadêmico, sem nenhuma barreira tecnológica ou financeira. Segundo a declaração “acesso aberto à literatura científica revisada por pares é a meta. O autoarquivamento e a nova geração de periódicos científicos de acesso aberto são as formas de se atingirem as metas”. Gomes e Rosa⁷ afirmam que este movimento foi o grande aliado nas mudanças no fluxo da comunicação científica.

Ao comemorar 10 e 15 anos do início do Movimento de Acesso Aberto, seus líderes renovam seus compromissos com o movimento e indicam novas recomendações. Nelas, seus idealizadores consideram, entre outras questões, as mudanças tecnológicas, que oferecem novas possibilidades em relação à produção e à disseminação da informação e do conhecimento científico.

Neste estudo, destacam-se, dentre as novas recomendações, aquelas que indicam a disponibilização de dados para serem utilizados como indicadores de métricas alternativas de impacto. Neste sentido, na próxima seção, serão apresentados os conceitos das principais métricas da informação, partindo daquelas consideradas tradicionais, até as novas métricas, voltadas para o ambiente da *Web*, destacando suas relações.

Métricas da informação

Faz parte dos estudos em Ciência da Informação compreender como se dão os processos de produção de conhecimento científico, o fluxo da informação para a sua produção e o desenvolvimento de indicadores que permitam avaliar sua qualidade, com o objetivo de entender como se dá sua evolução e auxiliar na tomada de decisões que envolvem o desenvolvimento da pesquisa em uma área do conhecimento. De forma mais ampla, a partir de elementos presentes na literatura científica, como a autoria, as citações, as palavras-chave, a afiliação e, também, baseado nas relações estabelecidas entre os atores que contribuem para o avanço da Ciência, estudiosos desenvolvem indicadores que permitem avaliar ou medir, entre outras coisas, a produtividade de autores e de instituições, sua popularidade e a de sua produção. Segundo Vanti⁸, “a avaliação, dentro de um determinado ramo do conhecimento, permite dignificar o saber quando métodos confiáveis e sistemáticos são utilizados para mostrar à sociedade como tal saber vem-se desenvolvendo e de que forma tem contribuído para resolver os problemas que se apresentam dentro de sua área de abrangência”⁸.

Assim foram sendo criadas as subdisciplinas Bibliometria, Cientometria, Patentometria, Informetria e, mais recentemente a Webometria e Almetria. Com enfoques diferentes, mas com funções semelhantes, essas subdisciplinas buscam, a partir de métodos quantitativos, medir a disseminação da informação e do conhecimento científicos, assim como, dar apoio à elaboração de políticas públicas voltadas ao desenvolvimento da Ciência.

A Bibliometria, segundo Guedes⁹ é uma área da Ciência da Informação, “constituída por um conjunto de leis e princípios empíricos, que contribuem para o estabelecimento de sua fundamentação teórica”⁹. Tague-Sutcliffe¹⁰ apresenta a bibliometria como “o estudo dos aspectos quantitativos dos processos de produção, disseminação e uso da informação registrada”¹⁰. O autor considera que o desenvolvimento de modelos e medidas, a partir da análise dos aspectos matemáticos estatísticos da produção científica, visa oferecer subsídios para os estudos de prospecção e de tomada de decisão, a partir da produção de indicadores.

Vanti⁸ destaca, dentro da disciplina Bibliometria, os autores que contribuíram na formulação das principais leis e princípios utilizados na área. São eles, Alfred James Lotka, George Kingsley Zipf e Samuel Clement Bradford. “A Lei de Lotka, ou Lei do Quadrado Inverso, aponta para a medição da produtivi-

dade dos autores, mediante um modelo de distribuição tamanho-frequência dos diversos autores em um conjunto de documentos. A Lei de Zipf, também conhecida como Lei do Mínimo Esforço, consiste em medir a frequência do aparecimento das palavras em vários textos, gerando uma lista ordenada de termos de uma determinada disciplina ou assunto. Já a Lei de Bradford, ou Lei de Dispersão, permite, mediante a medição da produtividade das revistas, estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas”⁸.

Aplicados à Biblioteconomia há, também, o estudo das citações, ou a análise de citações, que tem como principal destaque o *Science Citation Index (SCI)* do *Institute of Scientific Information (ISI)*, indicador produzido por Eugene Garfield, em 1961, a partir do qual foi possível mapear redes de citações em periódicos científicos. A análise de citações é uma das principais aplicações utilizadas nos estudos associados à produção científica, com a qual é possível observar, por exemplo, o impacto de uma publicação, a popularidade de uma pesquisa ou seu prestígio. Segundo Guedes⁹, a análise de citações “visa à mensuração da produção científica de um determinado pesquisador, de grupos de pesquisa, de instituições de pesquisa, países etc., como também a obtenção de indicadores acerca do desenvolvimento científico como um todo ou de um domínio específico do conhecimento”⁹.

Ainda associados à citação, existem duas análises denominadas acoplamento bibliográfico e cocitação. O acoplamento bibliográfico, segundo Guedes⁹, é “um procedimento de análise retrospectiva de citações que consiste na união de artigos citantes, citando o(s) mesmo(s) documento(s)”⁹. A cocitação, por sua vez, é “um procedimento de análise prospectiva de citações que consiste na união de artigos, citados pelo(s) mesmo(s) documento(s)”⁹. Ambas permitem identificar graus de ligação entre artigos ou autores, ora como citantes, ora como citados, permitindo, de certa forma, mapear redes de interesse e de produção de conhecimento.

Já a Cientometria, segundo Santos e Kobashi¹¹, “preocupa-se com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica”¹¹. Macias-Chapula¹² afirma que na Cientometria estão envolvidos estudos quantitativos relacionados às atividades científicas, a identifica como um segmento da sociologia da ciência e enfatiza o seu uso para o desenvolvimento de políticas científicas. Assim, segundo o autor, a Cientometria se situa “sobrepondo-se à bibliometria”¹². O autor também apresenta o conceito de Informetria que, de forma mais ampla, trata de aspectos quantitativos não somente da informação científica, mas da informação em qualquer formato, para os mais variados usos, não se limitando à comunidade científica. Assim, Macias-Chapula¹² afirma que “Informetria é o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social, e não apenas aos cientistas. A informetria pode incorporar, utilizar e ampliar os muitos estudos de avaliação da informação que estão fora dos limites tanto da bibliometria como da cienciometria”¹².

Nas últimas décadas, vem se constituindo um novo conjunto de métricas com o objetivo de avaliar os processos ligados à comunicação científica que passam a ser implementados utilizando-se da infraestrutura da Internet e, mais especificamente, os recursos da *Web* e da *Web 2.0*. Com os avanços tecnológicos, a literatura científica, suas fontes de consulta e seus processos de produção e disseminação vão migrando gradativamente para este novo ambiente, que passa a demandar de mecanismo adequados, assim como nos processos tradicionais, para estudar e avaliar seus fluxos de informação, as interações que acontecem entre seus atores e a qualidade, a popularidade e a visibilidade de suas produções, entre outras possibilidades.

Assim, surgem métricas alternativas ou complementares às métricas tradicionais já utilizadas nos estudos da Ciência da Informação, denominadas métricas da *Web*. Os principais exemplos são a Webmetria, a Webometria, a Cibermetria e a Altmatria.

A Webometria é caracterizada por Björneborn¹³ com um estudo quantitativo realizado a partir do uso de recursos de informação na *Web*, tendo o *link* como o principal elemento de estudo. O termo Webometria foi utilizado pela primeira vez por Almind e Ingwersen¹⁴, definido como os estudos informétricos aplicados à

Web. A partir do link, o qual Vanti¹⁵ faz correspondências às citações e às remissivas tão utilizadas em estudo da Biblioteconomia, estrutura-se uma rede de documentos. Assim, segundo a autora, “Ela [a *Web*] é constituída, basicamente, de sítios e *links* que corresponderiam, respectivamente, aos documentos e citações ou referências em um suporte impresso ou ainda aos itens ou registros e remissivas nos catálogos de bibliotecas tradicionais”¹⁵.

Neste universo, Thelwall, Vaughan e Björneborn¹⁶ apresentam quatro possibilidades de estudos baseados na análise do conteúdo das páginas *Web*; a análise da estrutura de seus *links*; a análise do uso do ambiente *Web*; e a análise das tecnologias para este ambiente. Um dos primeiros indicadores que surgiram associados à Webometria foi o fator de impacto na *Web* ou o *Web Impact Factor* (WIF), cuja definição é atribuída a Ingwersen¹⁷. Em sua primeira versão, eram considerados nos cálculos, número de páginas na *Web* (internas ou externas ao site) que possuíam pelo menos um *link* para o site em estudo. Este valor era dividido pelo número de páginas do *site*, tendo-se, assim, o fator de impacto do site de estudo. Outras versões foram propostas por Thelwall¹⁸ e Vanti, Costa e da Silva¹⁹, buscando diminuir inconsistências naturais dos processos de coleta de dados e problemas relacionados à manutenção e atualização das páginas, como afirmam Gouveia e Lang²⁰.

Gouveia²¹, por outro lado, apresenta alguns desafios que precisam ser enfrentados na realização de estudos Webométricos, destacando principalmente questões relacionadas à instabilidade e à indisponibilidade dos conteúdos, algumas vezes causados pela própria evolução tecnológica; as mudanças constantes nos mecanismos de buscas e nas ferramentas disponíveis para análise, eventualmente inviabilizando estudos de mais longo prazo.

De forma complementar, a Webmetria é um subconjunto da Webometria que “está associado a métricas de acesso de *sites* da *Web*, obtidas por análises de *logs* ou *page tagging*”²⁰. A Webmetria baseia-se no fato de que toda transação ou requisição executada na internet possui um registro e esses registros são armazenados em arquivos denominados arquivo de *log*. Assim, a partir desses dados, e utilizando as ferramentas de software adequadas, é possível, por exemplo, realizar estudos estatísticos sobre acessos a um determinado site, a origem desses acessos, visualizações, número de *downloads* de documentos, entre outras possibilidades. Para ampliar os dados referentes aos usuários que acessam conteúdo disponibilizado na Internet, estratégias como o *page tagging*, ou marcação de página, vêm sendo implementadas e oferecidas, gratuitamente ou em ferramenta de análise pagas, onde a mais popularmente conhecida ferramenta é o *Google Analytics*. Nessa estratégia é inserido um código executável pelo navegador, que se comunica com uma base de dados e deposita informações sobre o visitante virtual, como o tipo de navegador, resolução de tela, velocidade de conexão, e histórico de visitas ao sítio em análise, dentre outras informações, ampliando, assim, o conjunto de dados disponíveis para os estudos webométricos.

Outra área de estudos voltada para os fluxos de informação na *Web*, de forma mais ampla que a Webometria, é a Cibermetria ou Cybermetria. A Cibermetria abrange fluxos que estão além da *Web*, que é apenas uma das várias aplicações executadas na Internet. Assim, Björneborn²² a apresenta como o estudo dos aspectos quantitativos que consideram os fluxos informacionais que ocorrem em toda a Internet em aplicações como *e-mail*, *chats*, *mailing list*, *ftp* e a própria Internet. Gouveia e Lang²⁰ sintetizando os conceitos apresentados por Björneborn²², afirmam que “A cibermetria, portanto, é um campo mais amplo do que a Webometria, englobando a internet como um todo e não apenas a *Web*. Estudos sobre o uso da internet, de comunicações por e-mail, acesso a arquivos digitais, ambientes virtuais e redes de compartilhamento de arquivos seriam, assim, ciberométricos”²⁰.

Por seu turno, a altmetria, mais recentemente, e com crescimento acelerado, vem contribuindo para o estudo dos novos fenômenos observados na *Web*, relacionados com interações entre os usuários e os produtores de informação na internet. Com o uso crescente das redes sociais na *Web* pelos membros da comunidade acadêmica, observa-se que este canal também pode oferecer indicadores que apontem novos processos dentro do sistema da comunicação científica. Priem, autor do termo *altmetrics*, o utilizou pela primeira vez em um ambiente de rede social, em um *post* no *Twitter*. Posteriormente, Priem e colabora-

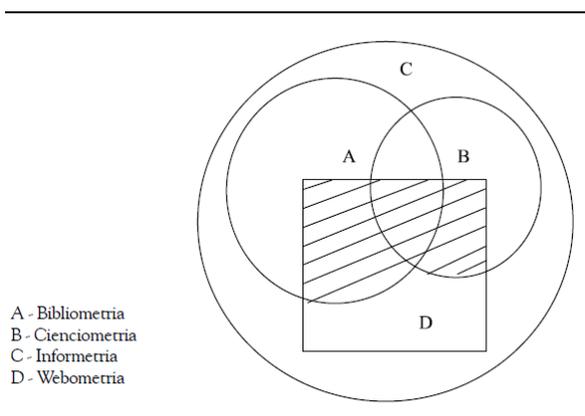
dores²³ o consolidaram no texto *Altmetrics: a manifesto*, indicando a possibilidade de serem observados novos processo de disseminação da produção científica na Web Social.

Com o surgimento desses novos campos do conhecimento dedicados a análises quantitativas dentro da Biblioteconomia e da Ciência da Informação, mas que possuem interseções com os campos tradicionais dessas áreas do conhecimento, vários autores buscaram esclarecer e representar seus limites e suas inter-relações, considerando para cada caso, sua abrangência, seus recursos e seus objetos de estudos^{8,15}. É importante frisar que os modelos apresentados a seguir representam as reflexões dos autores sobre os novos campos em relação aos tradicionais. Estes modelos que vem sendo amadurecidos, ao longo dos últimos 20 anos, desde que as métricas para a Web passaram a ser tema de estudo pela comunidade acadêmica.

Vanti⁸, ao abordar os conceitos de bibliometria, cientometria, informetria e webometria, e considerando as características de cada subcampo apresentadas por Macias-Chapula¹², elaborou o diagrama a seguir, representado na Figura 1.

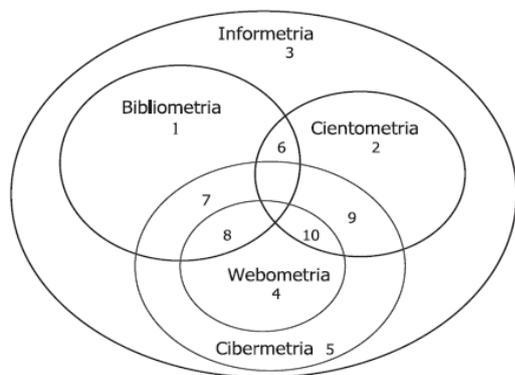
Em uma segunda abordagem, apoiando-se nas considerações de Thelwall Vaughan e Björneborn¹⁶, Björneborn e Ingwersen²⁴ e Björneborn¹³ sobre as características de cada uma dessas métricas, Vanti¹⁵ apresentou novo diagrama, no qual foi incluída a cibermetria, e esta englobando a webometria, uma vez que seu objeto de estudo está representado por toda a internet, incluindo toda a Web. A Figura 2 mostra esta nova representação.

Figura 1 – Diagrama da inter-relação entre os campos Bibliometria, Cientometria, Informetria e Webometria



Fonte: VANTI, 2002

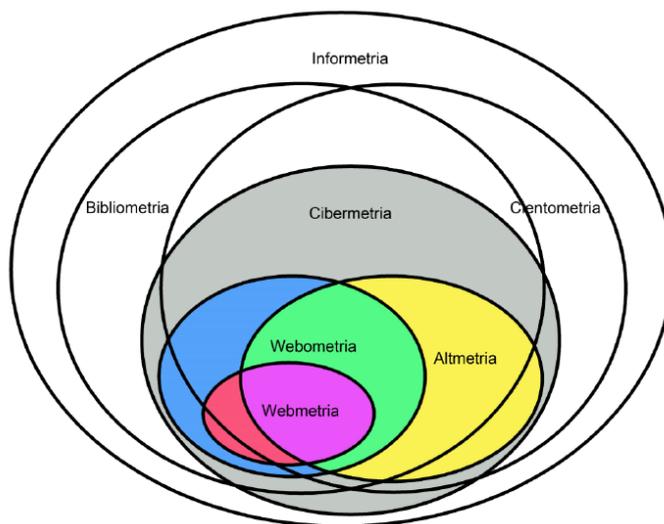
Figura 2 – Diagrama da inter-relação entre os campos Bibliometria, Cientometria, Informetria e Webometria



Fonte: VANTI, 2005

Mais recentemente, Gouveia²⁵ elabora novo diagrama, onde são acrescentadas ao grupo de métricas a webmetria e altmetria. É possível observar, como já foi apresentado neste trabalho, que a webmetria encontra-se totalmente inserida no conjunto da webometria, pelo fato de considerar outros elementos dentro do universo da *Web*, além do *link*. Já a altmetria, apesar do manifesto altmétrico apresentá-la como um campo completamente desassociado da webometria ou da cibermetria, Gouveia²⁵ entende que a altmetria utiliza dados, tanto da Webometria, quanto da Cibermetria para realizar estudos cientométricos. Assim, a Figura 3 apresenta essa nova representação.

Figura 3 – Diagrama da inter-relação entre os campos Altmetria, Bibliometria, Cibermetria, Cientometria, Informetria, Webmetria e Webometria



Fonte: GOUVEIA, 2013

No Brasil, segundo dados do diretório *OpenDoar*, os artigos de periódicos revisados por pares e as teses e dissertações são os principais documentos depositados nos RI de acesso aberto. Uma vez que os repositórios institucionais passam a fazer parte do fluxo da comunicação científica, facilitando o acesso à produção intelectual de uma instituição, os dados relativos às interações ocorridas no ambiente da *Web*, em relação a esse sistema, podem fornecer informações relevantes sobre a produção e o uso da informação científica.

Shintaku e colaboradores²⁶ afirmam que a representatividade dos repositórios para a comunicação científica necessita de avaliação baseada em indicadores. Os autores complementam, indicando que esse novo formato de produção, disseminação e acesso à informação científica, por meio dos RI de acesso aberto, carece de avaliação baseada em indicadores, assim como o sistema tradicional de comunicação científica. Sendo um recurso disponível no ambiente da *Web*, os autores sugerem que a webometria se configura com uma ferramenta adequada para esse tipo de análise.

A construção do endereço de acesso ao repositório, ou a sua URL – *Uniform Resource Locator* – também se caracteriza como um elemento importante nos estudos webométricos, como indica o projeto *Cybermetric Lab*ⁱⁱⁱ, responsável pelo *Ranking Web of World Repositories*^{iv}. Não só por ser uma forma de identificar a autoria institucional de uma determinada produção, mas, também, pelo fato de que a forma como esses URL são construídos pode facilitar ou dificultar a análise por meio de ferramentas específicas de coleta e análise de dados webométricos.

iii Projeto de pesquisa do Consejo Superior de Investigaciones Científicas – CSIC – (<http://repositories.webometrics.info/es/node/18>)

iv <http://repositories.webometrics.info/es>

O *Uniform Resource Locator* é uma indicação – um endereço – para a localização de uma informação na *Web*. Ele é composto por partes nas quais são indicados o tipo de protocolo utilizado (http, ftp etc.), seguido do nome de domínio do servidor onde a informação está hospedada e o caminho para esta informação (um arquivo) no servidor. Já o nome de domínio é um esquema composto por nomes separados por pontos que visam representar sites ou grupos ou a autoridade responsável pelo servidor onde a informação está disponível para acesso. Como exemplo de um URL temos, <http://www.instituicao.br>, onde http é o protocolo utilizado e www.instituicao.br representa a autoridade responsável pelas informações²⁷.

Orduña-Malea e López-Cózar²⁸, colaborando com os padrões indicados pelo *Cybermetric.Lab*, ressaltam a importância do uso adequado dos nomes de domínio – parte integrante do URL – para acesso aos repositórios e para a sua indexação por mecanismos de busca. Os autores chamam a atenção para o fato de que “os repositórios institucionais são, em última instância, *websites*”²⁸ e como tais devem considerar em sua estrutura, além das questões como usabilidade e arquitetura da informação, também a otimização para sua indexação e recuperação por mecanismos de busca na internet. Observando os dados apresentados pelo diretório *OpenDOAR* e pelo *ranking* de repositórios do *Cybermetric Lab*, relativos ao RI da América Latina, Orduña-Malea e López-Cózar²⁸ verificam que o *ranking* contempla pouco mais de 50% dos RI cadastrados no *OpenDoar*. Para os autores, o uso inadequado dos nomes de domínio dos RI é o principal fator responsável por este fenômeno.

O *Cybermetric Lab* recomenda um padrão para a construção da URL dos RI de forma que no nome de domínio estejam presentes o nome do repositório e a sigla da instituição, como, por exemplo, *repositorio.usp.br* ou *repositorio.ufrb.edu.br* ou *repository.ucla.edu*.

Metodologia

Trata de uma pesquisa descritiva exploratória, tendo como fonte os *sites* dos RI de IPES brasileiras, onde são observados elementos que possam se constituir em indicadores para análise webométrica e altmétricas da produção científica neles depositados.

Braga²⁹ nos mostra que a pesquisa exploratória se caracteriza pela reunião de “dados, informações, padrões, ideias ou hipóteses sobre um problema”²⁹, assim como a pesquisa descritiva tem como objetivo a identificação de características de um determinado problema. Desta forma, acessando o *site* de cada RI das IPES brasileiras, foram coletados dados e informações que permitiram observar como esse ambiente vem se adaptando às novas modalidades de indicadores e análise de visibilidade, prestígio, citações, entre outros, na *Web*.

A pesquisa foi realizada por meio do acesso aos *sites* dos RI das IPES, tendo como base a relação disponível no diretório *OpenDOAR*^v. O acesso a todos os sites foi realizado no dia 20 de abril de 2017. Nestes repositórios, foram observados o padrão de construção dos URL – *Uniform Resource Identifier* –; o uso de ferramentas que geram estatísticas relacionadas ao acesso ao *site*, ao objeto digital e *downloads* realizados; o uso de indicadores de redes sociais, como *Facebook*, *Twitter*, *Google+*, entre outras; e o uso de identificadores únicos de objetos digitais.

A partir dos dados obtidos nos sites foram elencados os percentuais que correspondem a ocorrências dos elementos pesquisados, no conjunto dos RI estudados, apresentando o cenário dos repositórios institucionais de acesso aberto em relação a dados *webométricos*, *webmétricos* e *altmétricos*.

v The Directory of Open Access Repositories – OpenDOAR (<http://www.openoar.org/>)

Resultados: Apresentação e discussão

A coleta de dados foi realizada no dia 20 de abril de 2017, inicialmente no *site* de 33 RI. Neste dia, 3 *sites* apresentaram problemas devido à indisponibilidade do serviço ou algum outro tipo de problema que impediu a coleta de forma satisfatória de todos os dados necessários para a pesquisa. Desta forma, os dados analisados correspondem aos 30 RI de IPES brasileiras.

Observou-se a construção das URL dos repositórios, seguindo o padrão orientado pelo projeto *Cybermetric Lab*^{vi}, responsável pelo *Ranking Web of World Repositories*^{vii}, que indica que na URL estejam presentes o nome do repositório e a sigla da instituição, como, por exemplo, *repositorio.usp.br* ou *repositorio.ufrb.edu.br* ou *repository.ucla.edu*.

Com relação aos RI pesquisados, vimos que 21, ou seja, 70% do total de RI, apresentam seu URL no formato recomendado. Nos demais casos há outros elementos que compõem o endereço de acesso ao repositório e, nestes casos, serão necessários estudos mais específicos para identificar as principais dificuldades que estes formatos possam ocasionar na obtenção de dados por meio de ferramentas de análise.

Observou-se, ainda, que 24 RI das instituições disponibilizam informações estatísticas sobre seus objetos digitais, ou seja, 80% do total analisado. Neste caso, são disponibilizadas informações sobre número de acessos ao item, número de *downloads*, origem (regiões do país ou do mundo) desses acessos. A partir desses dados é possível realizar, por exemplo, estudos que permitam identificar temas de pesquisa de maior interesse, o prestígio de determinado pesquisador ou de determinada instituição.

Com relação aos indicadores altmétricos, como os estudos têm demonstrado, ainda são utilizados em pequena escala. Nos sites dos RI pesquisados foi possível observar que ainda são poucos os que disponibilizam dados ou mecanismos que possibilitem o estudo de novos processos de disseminação da produção científica na Web Social e das interações entre usuários e produtores de informação científica na internet. Dos 30 sites de RI observados, 9, ou seja, 30% do total, disponibilizam algum recurso social com indicadores altmétricos, com maior ocorrência das redes sociais *Facebook* e *Twitter*.

Os resultados também mostram que 28 instituições, ou seja, cerca de 93% do total, utilizam sistema persistente de identificação de objeto digital, fator importante para os estudos métricos na web, uma vez que identifica unicamente o objeto digital e, além disso, busca garantir a preservação do acesso ao item, independente de sua localização na internet. No caso deste estudo, 28 dos 30 RI pesquisados que utilizam sistema persistente de identificação de objeto digital também utilizam o sistema *DSPACE*. O sistema de identificação persistente utilizado é o *Handle System*, que acompanha o pacote de *software* do *DSPACE*. Certamente este fato favorece a utilização desse sistema pelos RI estudados.

Considerações finais

O presente estudo teve como objetivo observar como os RI de acesso aberto das IPES estão se adequando aos novos métodos de análise dos fluxos de informação na comunicação científica, frente às novas métricas da web. Com a intensificação do uso da internet e mais especificamente a web nos processos de comunicação científica, torna-se necessário, que os sistemas de gestão desse conteúdo, como os RI neste estudo, se adéquem a esses novos mecanismos de análise webométricas, webmétricas e altmétricas, disponibilizando dados que permitam a realização de estudos alternativos e complementares aos tradicionais, já desenvolvidos na Ciência da Informação, com foco nas citações, prestígio, visibilidade, entre outros.

vi Projeto de pesquisa do *Consejo Superior de Investigaciones Científicas* – CSIC – (<http://repositories.webometrics.info>)

vii <http://repositories.webometrics.info/es>

Nesse estudo, foi possível observar que os RI das IPES brasileiras estão acompanhando esta evolução, buscando adequar seus *sites* de forma que favoreçam análises webométrica e altmétrica pelas principais ferramentas disponíveis na web. Nota-se que há maior ênfase no uso de recursos associados à produção de dados webmétricos, representado por 70% dos *sites* que disponibilizam informações sobre acesso e *download* de seu conteúdo. Mas ainda é incipiente o uso de recursos associados à altmetria, representado por 22% dos *sites* que disponibilizam dados relacionados a interações em redes sociais. Fica evidente, também a preocupação em relação à identificação única da produção institucional e à preservação de seu acesso, independente de questões relacionadas ao seu endereço de localização na internet, onde observamos que a grande maioria dos RI, 83%, utilizam o sistema *Handle*, serviço nativo nas versões atuais do DSPACE, que é o sistema utilizado em 30, dos 31 RI pesquisados.

Assim, os resultados mostram que os *sites* dos RI das IPES brasileiras estão se adequando de forma que favoreçam as análises webométrica e altmétrica, tendo este processo de adequação facilitado pelo uso do sistema DSPACE, que oferece recursos que contribuem para este tipo de análise. Importante ressaltar que as iniciativas do *Cybermetric Lab* também têm contribuído nesse cenário, estimulando o estabelecimento de padrões para construção dos endereços de acesso aos *sites* dos repositórios e para a filosofia de acesso aberto às publicações científicas.

Referências

1. SAYÃO, L., et al. Implantação e Gestão de Repositórios Institucionais: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009.
2. CROW, R. The Case for institutional repositories: ASPARC position paper. Washington, DC: Scholarly Publishing & Academic Resources Coalition, 2002. Disponível em: <<http://www.arl.org/sparc/IR/ir.htm>>.
3. LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. ARL Bimonthly Report, 26, 2003. Disponível em: <<http://www.arl.org/newsltr/226/ir.html>>.
4. LEITE, Fernando César Lima. Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto. Brasília: IBICT; 2009.
5. RODRIGUES, E.; SWAN, A.; BAPTISTA, A.A. Uma década de acesso aberto na UMinho e no mundo. Porto: Universidade do Minho; 2013. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/26144/3/RepositoriUM_10anos.pdf>.
6. BOAI, Budapest Open Archives Initiative. 2002. Disponível em: <<http://www.budapestopenaccessinitiative.org/read>>.
7. GOMES, Maria João; ROSA, Flávia. Repositórios Institucionais: democratizando o acesso ao conhecimento. Salvador: UFBA, 2010.
8. VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. Ciência da Informação, v. 31, n. 2, p.152-162, 2002. Disponível em: <<http://revista.ibict.br/ciinf/article/view/970/1007>>.
9. GUEDES, V.L.S. A bibliometria e a gestão da informação e do conhecimento científico e tecnológico: uma revisão da literatura. Ponto de acesso, v.6, n.2, 2012. Disponível em: <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/5695>.
10. TAGUE-SUTCLIFFE, J. An introduction to informetrics. Information Processing & Management, v. 28, issue 1, p.1-3, Jan./Feb. 1992. Disponível em: <<http://www.sciencedirect.com.ez29.periodicos.capes.gov.br/science/journal/03064573/28/1>>.
11. SANTOS, R. N. M.; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. Tendências da Pesquisa Brasileira em Ciência da Informação. Brasília, v. 2, n. 1, p. 155-172, jan./dez. 2009. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/21/43>
12. MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. Ciência da Informação, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-19651998000200005&script=sci_abstract&tng=pt>.

13. BJÖRNEBORN, Lennart. Small-world link structures across an academic web space: a library and information science approach, 2004. Royal School of Library and Information Science. Tese de Doutorado. 469p. Disponível em: <http://forskning.ku.dk/find-en-forsker/?pure=files%2F47039808%2Fennart_bjorneborn_phd.pdf>.
14. ALMIND, T.C.; INGWERSEN, P. Informetric analyses on the world wide web: methodological approaches to 'Webometrics'. Journal of Documentation, v. 53, n. 4, p. 404-426, 1997.
15. VANTI, N. Os links e os estudos webométricos. Ciência da Informação, v. 34, n. 1, p. 78-88, 2005. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1104/1226>>.
16. THELWALL; VAUGHAN, L.; BJÖRNEBORN, L. Webometrics. Annual Review of Information Science And Technology, v. 39, 2003. Preprint.
17. INGWERSEN, P. The calculation of web impact factors. Journal of Documentation, v. 54, n. 2, p. 236-243, 1998.
18. THELWALL. Sources of links for WIF calculations. Journal of Documentation, v. 58, n. 1, p. 60-72, 2002.
19. VANTI, N.; COSTA, J. A .F.; DA SILVA, I. C. O. Nova fórmula revisada para o cálculo do fator de impacto web (FIW). Liinc em revista, v. 9, n. 1, p. 228-236, 2013. Disponível em: <<http://revista.ibict.br/liinc/article/view/3421/2996>>.
20. GOUVEIA, Fábio Castro; LANG, Pamela. Da webometria à altmetria: uma jornada por uma ciência emergente. In: ALBAGLI, Sarita (Org). Fronteiras da Ciência da Informação. Brasília: IBICT; 2013.
21. GOUVEIA, F. C. Novos caminhos e alternativas para a Webometria. Em questão (UFRGS. Impresso), v. 18, p. 249-261, 2012. Disponível em: <<http://seer.ufrgs.br/EmQuestao/article/viewFile/33270/23785>>.
22. BJÖRNEBORN, L. Small-world link structures on the web. Copenhagen, DK: School of Library and Information Science, 2002. Disponível em: <www.db.dk/lb/2002smallworld.pps>.
23. PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. Altmetrics: A manifesto, (v.1.0), 26 Outubro 2010. Disponível em: <<http://altmetrics.org/manifesto>>.
24. BJÖRNEBORN, L.; INGWERSEN; P. Towards a basic framework of webometrics. Journal of American Society for Information Science and Technology, v. 55, n. 14, 2004, p. 1216-1227.
25. GOUVEIA, F. C. Altmetria: métricas de produção científica para além das citações. Liinc em revista, v. 9, n. 1, p. 214-227, 2013. Disponível em: <<http://revista.ibict.br/liinc/article/view/3434>>.
26. SHINTAKU, M., ROBREDO, J., BAPTISTA, D.M., Webometria dos repositórios institucionais acadêmicos. Ciência da Informação. DF, v. 40 n. 2, p. 312-326, maio/ago. 2011. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1929/1428>>.
27. Comer, D.E. Internetworking with TCP/IP: principles, protocols, and architecture. 3a Edição. Prentic-Hall, New Jersey, 1995.
28. Orduña-Malea, López-Cózar. Low visibility of Latin American repositories in Google Scholar: technical incompatibility or lack of web strategy? LSE Impact Blog. 2014. Disponível em: <http://blogs.lse.ac.uk/impactofsocialsciences/2014/07/31/latin-american-repositories-google-scholar-low-visibility/>.
29. BRAGA, Kátia Soares. Aspectos relevantes para a seleção de metodologia adequada à pesquisa social em Ciência da Informação. In: MULLER, Suzana Pinheiro Machado. Métodos para a pesquisa em Ciência da Informação. Brasília: Thesaurus; 2007.