

XIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB 2013)
GT 8: Informação e Tecnologia

Comunicação Oral

**PUBLICAÇÕES AMPLIADAS: UM NOVO MODELO DE PUBLICAÇÃO
ACADÊMICA PARA O AMBIENTE DE E-SCIENCE**

Luana Farias Sales – IBICT/UFRJ

Luís Fernando Sayão – IBICT

Rosali Fernandez Souza – IBICT

Resumo

A pesquisa científica atual se desenrola sob as condições de um paradigma científico emergente, conhecido como e-Science ou quarto paradigma. Essa nova forma de fazer ciência se distingue pelo uso intensivo de redes de computadores, repositórios digitais distribuídos e pela geração extraordinária de dados de pesquisa em formatos digitais. O objetivo deste estudo é analisar como os dados de pesquisa aliados às tecnologias de informação podem enriquecer as publicações tradicionais criando novas formulações de documentos científicos que comportem a complexidade das pesquisas científicas do paradigma atual, rompendo a barreira do texto e criando relações semânticas entre *e-prints* e dados e coleções de dados de pesquisa. Nessa direção, a metodologia deste trabalho enfoca o estudo e análise da estrutura, padrões e tecnologias do modelo de publicação científica chamado de publicações ampliadas. O resultado deste estudo e análise é o que é apresentado neste artigo.

Palavras-chave: Publicações Ampliadas. Dados Digitais de Pesquisa. Norma Oai-Ore. E-Science.

**ENHANCED PUBLICATIONS: A NEW MODEL OF SCHOLARLY PUBLICATION
FOR THE ENVIRONMENT OF E-SCIENCE**

Abstract: The scientific research takes place under the conditions of an arising paradigm, known as e-Science or fourth Paradigm. This new way of perform science is characterized by intensive application of computer networks, distributed digital repositories and by extraordinary generation of research data in digital formats. The objective of the study is to analyze how research data combined with information technologies can enrich traditional publications creating new concepts of scientific documents that can deal with the complexity of scientific research of the current paradigm, breaking the barrier of text and creating semantic relationships between e-prints and data and data collections of research data. In this direction, the methodology of this work focuses on the studies and analysis of the structure, standards and technologies of scientific publication model called enhanced publications. The result of this study and analysis is what is presented in this article.

Keyword: Enhanced Publications. Digital Data Research. Oai-Ore. E-Science.

1 INTRODUÇÃO E MOTIVAÇÃO

A comunidade científica vem sentindo a necessidade de novos modelos de publicações que possam expressar a complexidade e a dinâmica da pesquisa científica contemporânea. Os pesquisadores demandam por modelos que explicitem mais claramente a pesquisa desenvolvida, seus métodos, seu material, os dados e conhecimentos gerados; que revelem com clareza a intenção do pesquisador; e que os resultados sejam avaliados de forma mais eficaz e interativa. A pesquisa científica atual necessita de publicações que expandam as formas de socialização e de trocas e que sejam capazes de ajudar na replicação de experimentos e em testes de veracidade ou ofereçam seus conteúdos para reutilização em outros contextos. Os padrões convencionais, mesmo digitais, que oferecem apenas uma mímica eletrônica da versão impressa, como é o documento em formato PDF, não permitem o usuário interagir com outros pesquisadores e também com o ambiente tecnológico que o envolve, condições cada vez mais críticas para a pesquisa, já que atualmente a ciência se faz de forma cada vez mais colaborativa e interdisciplinar.

Não obstante todas as transformações comportamentais e sociais decorrentes do aparato tecnológico que permeia e transforma o universo atual da pesquisa (por exemplo, a geração, o uso, reuso e compartilhamento intensivo de dados em formatos digitais) a infraestrutura atual de comunicação científica ainda está fortemente centrada no armazenamento e na disseminação de recursos informacionais individuais. Partindo dos modelos de publicação na *Web* e voltando aos sistemas formais de informação acadêmica, como as bibliotecas de pesquisa, verifica-se que eles entregam ao usuário basicamente um artigo ou uma monografia. A própria indústria editorial acadêmica está fortemente ancorada neste modelo e não aceita os produtos da e-pesquisa¹, como base de dados e simulações, por exemplo. O que parece cada vez mais claro é que a heterogeneidade e a complexidade dos registros de resultados de pesquisa não podem mais ser expressas por documentos convencionais únicos, impressos ou mesmo digitais (SAYÃO; SALES, 2012).

E-pesquisa ou e-research como é chamado na literatura internacional é a pesquisa científica realizada com o apoio da tecnologia de informação, entre suas características principais pode-se destacar a geração intensiva de dados principalmente por meio do uso de computação em grade (grid technology).

Mas esta é uma inquietação antiga. No início do século passado, ainda muito distante da revolução tecnológica que marca o tempo presente, Paul Otlet, pai da Documentação, já criticava o modelo tradicional de publicação e, em diversas de suas palestras já mencionava a necessidade de um novo modelo, levantando a possibilidade de criação de um novo formato de livro que deveria ter por base uma forma colaborativa de representação do conhecimento. (VAN DEN HEUVEL e RAYWARD, 2011).

Desde meados da década de 1990, o advento da Internet vem trazendo diversas oportunidades de melhoria dos canais de comunicação científica, por exemplo as publicações *online* revolucionaram as possibilidades de descoberta e de acesso a recursos informacionais acadêmicos e científicos, e inauguraram possibilidades alternativas de publicação, aumentando a amplitude e a facilidade de acesso a todo tipo de material, inclusive conjuntos de dados científicos. No entanto, quando se trata do artigo científico - enquanto principal veículo para comunicar pesquisa - as publicações apresentaram pouca melhoria, principalmente no que diz respeito à aplicação de tecnologias de Web 2.0 e Web Semântica, que poderiam certamente beneficiar esse veículo em termos de representação, estrutura, organização, apresentação e interatividade. Essa constatação pode ser evidenciada no primeiro texto de Kircz (2001) sobre novas práticas para publicação eletrônica, onde o autor mostra que não basta juntar informações, mas é preciso reuni-las sob um contexto para que possam servir de base para a argumentação entre os pares.

“um documento eletrônico não é uma versão eletrônica de um documento tradicional em papel, com enfeites, como *hiperlinks*, fotos coloridas e animações ilustrativas. Em vez disso, um documento eletrônico é um documento que compreende uma variedade de diferentes tipos de informações que são apresentadas juntas por um autor, a fim de formular um argumento científico abrangente.” (KIRCZ, 2001, p.266)

Segundo Aalbersberg, Dunham e Koers (2011, p.2), nas últimas décadas, as “melhorias no artigo acadêmico têm sido principalmente em termos de entrega (eletronicamente), descoberta (pesquisa de texto completo), bem como uma série de melhorias específicas, em menor escala, como a possibilidade de fazer *upload* de dados complementares”. Entretanto, em termos de estrutura e forma, o artigo atual é, em geral, o mesmo que nas primeiras revistas científicas do século XVII. É então com o intuito de estabelecer uma nova infraestrutura para disseminação e compartilhamento do conhecimento científico que novos modelos de publicação, que utilizam o máximo potencial tecnológico, vêm sendo propostos.

Recentemente vários estudos se concentraram na possibilidade de se entrelaçar produtos de e-pesquisa que se encontram distribuídos, expandindo o conceito tradicional de documento científico (AALBERSBERG, DUNHAM E KOERS, 2001¹, KIRCZ, 2001, 2002; VEERNOOY-GERRITSEN, 2009). Diversos modelos são encontrados na literatura - documentos sobrepostos², documentos compostos³, publicações semânticas⁴, publicações ampliadas - dentre eles destaca-se o último citado por reunir entre suas características, elementos de todos os outros tipos e possuir um grau de formalização tecnológica e informacional mais elaborado, e ter ainda um maior grau de aceitação pelos principais atores envolvidos: autores, instituições de pesquisa, agências financiadoras de pesquisa e editores. É este modelo que será apresentado no presente trabalho.

O presente estudo é parte de uma pesquisa em andamento, cujo objetivo é investigar interdisciplinarmente os novos conceitos de publicação científica que emergem no ambiente estabelecido pelos postulados da chamada *e-science* ou quarto paradigma científico. A pesquisa considera coletivamente os impactos destas publicações na comunicação científica, as questões relativas à curadoria de dados digitais de pesquisa e, com maior ênfase, a proposta de um modelo semântico que relaciona dados de pesquisa com *e-prints* – publicações acadêmicas tradicionais em formatos digitais, tais como artigos e teses. O campo empírico onde a pesquisa se desenrola é na área de ciências nucleares, mais especificamente no Instituto de Engenharia Nuclear e seu repositório digital de dados e informações. O foco do presente trabalho é uma análise dos conceitos, tecnologias e padrões que envolvem as publicações ampliadas.

2 PUBLICAÇÕES AMPLIADAS: DEFINIÇÃO, ESTRUTURA, PADRÕES, COMPONENTES

A atividade de pesquisa científica do século XXI produz uma quantidade extraordinária de dados, principalmente em formatos digitais. Isto acontece essencialmente porque a tecnologia digital se torna mais e mais um elemento onipresente nos processos da construção do conhecimento científico, principalmente no desenvolvimento de instrumentos

² Documentos sobrepostos - uma publicação que não publica nenhum artigo original, mas seleciona artigos existentes em outros lugares, acrescenta certo valor e publica os resultados como um serviço para os usuários de sua base (VAN DE SOMPEL et al, 2006).

³ Documentos compostos – modelo de documento que encapsula vários *datasets* e recursos gerados ou utilizados durante a realização de um experimento científico, ou descoberta do processo, dentro de um único objeto composto para publicação e intercâmbio. (CHEUNG, LASHTABEG E DRENNAN, 2008)

⁴ Documentos semânticos – documentos que são combinados com ontologias permitindo que usuários acessem o conhecimento de várias maneiras.(ERICKSON, 2007).

científicos, na aplicação intensiva de tecnologias computacionais de alto desempenho por meio de simulações e no trabalho colaborativo em escala planetária por meio dos laboratórios⁵. Percebe-se que na prática, o problema crítico da ciência hoje não é escassez, mas sim excesso de dados. Como desdobramento desse fato, torna-se crítico o desenvolvimento de metodologias - tecnológicas e gerenciais - para a geração e formação de coleções de dados, arquivamento, preservação, análise e interpretação desses dados numa diversidade transversal de contextos.

Com o reconhecimento da importância para as “ciberinfraestruturas” mundiais de pesquisa do arquivamento, gestão e curadoria dos dados para posterior acesso e reuso em outros ambientes e também para validação e replicação das pesquisas, vem crescendo o número de repositórios confiáveis que armazenam dados científicos, chamados repositórios digitais de dados de pesquisa.

Este fato somado à “constatação de que a publicação tradicional é limitada na sua capacidade para incorporar os resultados de todo o processo de descoberta científica, especialmente quando grandes conjuntos de dados são gerados” (VERHAAR, 2007, p.11), coloca a possibilidade de estabelecimento de um novo modelo de publicação que pode atender as expectativas da comunidade acadêmica e científica, isto é um modelo que possa ligar *e-prints*⁶ e dados científicos de forma dinâmica, expandindo a fronteira das publicações tradicionais.

Em 2007, o Digital Repository Infrastructure Vision for European Research - II (DRIVER-II), desenvolvido pela Comunidade Europeia, publicou um relatório sobre modelo de objetos digitais onde definiu o termo *enhanced publication* ou publicação ampliada - como estamos convencendo chamar no país - da seguinte forma: “objetos digitais compostos que combinam *e-prints* com um ou mais recursos de dados, um ou mais registros de metadados, ou qualquer combinação destes”.

⁵ Colaboratório é um termo cunhado por Wulf (1989) para designar um “centro de pesquisa sem paredes, no qual os pesquisadores de um país podem realizar suas pesquisas sem considerar a localização física, interagindo com colegas, acessando instrumentação, compartilhando dados, informações e recursos computacionais, e acessando bibliotecas digitais”

⁶ *E-prints* são recursos eletrônicos textuais que expressam um trabalho acadêmico original, destinado a ser lido por seres humanos, que apresenta algumas reivindicações acadêmicas e que geralmente contém uma interpretação ou uma análise de determinados dados primários. (VERHAAR, 2007)

De uma forma simples, uma publicação pode ser ampliada a partir da agregação de um ou mais recursos a um *e-print*. Estes recursos podem ser dados de toda a natureza, outros *e-prints* e metadados e podem ser ainda recursos produzidos ou consultados durante a criação do texto e que, geralmente, apoiam, justificam, ilustram ou esclarecem as afirmações científicas que são apresentadas em uma publicação. Woutersen-Windhouver e Brandsma (2009, p.136) afirmam que uma “publicação é ampliada com dados de pesquisa, materiais extras, dados de pós-publicação, registros de base de dados (ex. *protein data bank*) e tem uma estrutura baseada em objeto com *links* explícitos entre objetos”. Assim, um objeto pode ser parte de um artigo, um *data set*, uma imagem, um filme, um comentário, um módulo ou um *link* para informação em uma base de dados.

De forma geral, um manuscrito acadêmico é arquivado em um repositório digital institucional ou temático – ou seja, um repositório de *e-prints*; outros componentes, arquivados no mesmo ou em outros repositórios, são agregados ao manuscrito como parte do fluxo de trabalho do desenvolvimento da atividade de pesquisa científica. Um exemplo simples de uma publicação ampliada pode consistir num *e-print* combinado meramente com um conjunto mínimo de metadados. A figura 1 representa uma generalização do processo de ampliação, considerando os dados variando no tempo, indicando a possibilidade de versionamento da publicação e de seus componentes.

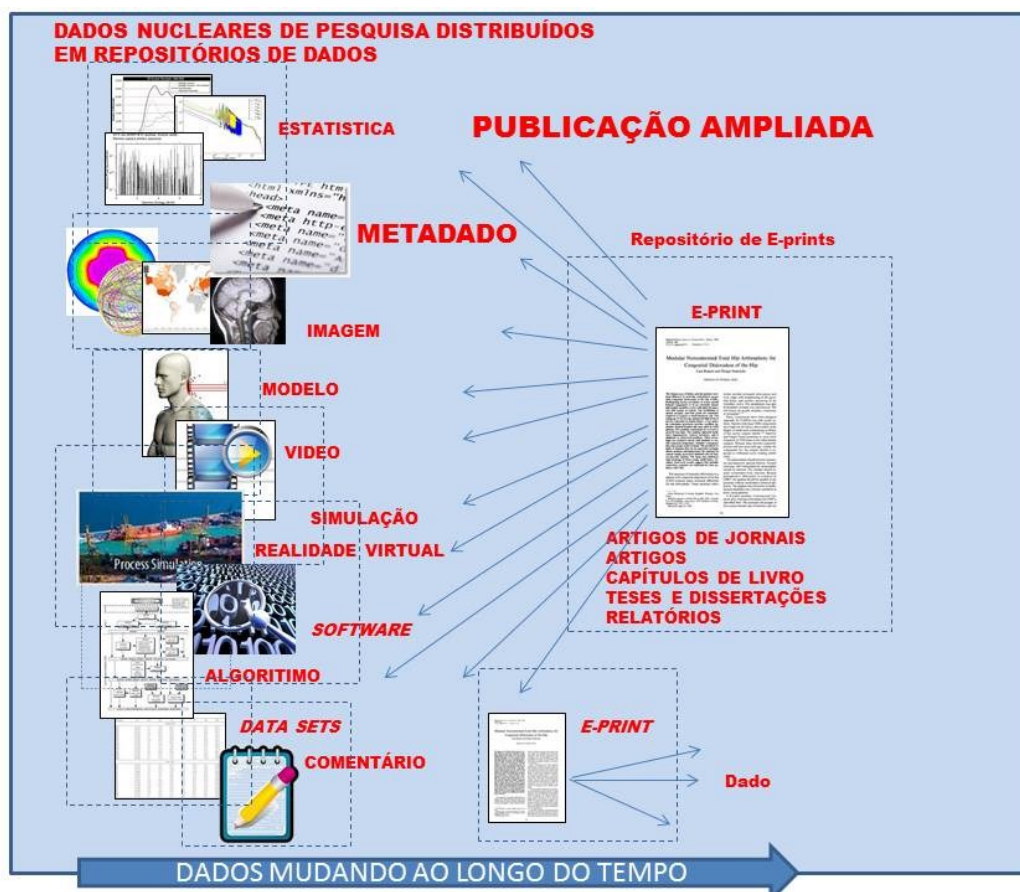


Figura 1: E-prints ampliados por dados de pesquisa
 Fonte: O autor

No relatório de Vehaar (2007) foi identificada ainda uma série de dez requisitos e recomendações que fundamentam uma publicação ampliada. Esses itens estabelecem uma base para a formalização tecnológica e informacional dessas publicações, aumentando as possibilidades de recuperação, acesso, uso e reuso, interoperabilidade e preservação. São os seguintes:

- a) deve ser sempre possível especificar as partes componentes de uma publicação ampliada;
- b) a publicação ampliada e seus componentes devem ser disponibilizados como recurso *web* identificados por URI⁷, o mesmo vale para seus componentes;
- c) deve ser possível agregar objetos digitais compostos à publicação ampliada;
- d) deve ser possível acompanhar as diferentes versões das publicações ampliadas como um todo e das suas partes constituintes;

⁷ Uniform Resource Identifiers

- e) deve ser possível registrar as propriedades básicas da publicação e dos outros recursos que estão adicionados a ela;
- f) deve ser possível registrar a autoria da publicação ampliada e dos outros recursos que a compõe;
- g) deve ser possível assegurar a preservação de longo prazo das publicações ampliadas;
- h) deve ser possível registrar as relações entre os recursos *web* que fazem parte da publicação ampliada;
- i) As instituições que oferecem acesso a publicações ampliadas devem assegurar que elas possam ser descobertas;

As instituições que oferecem acesso a publicações ampliadas devem assegurar que elas estejam disponíveis como documentos baseados no modelo OAI-ORE (que será discutido mais adiante).

Estes requisitos são apresentados na figura 2, expresso na forma de um diagrama entidade-relacionamento. O modelo mostra que as publicações ampliadas consistem de cinco tipos de entidades: *e-prints*, dados, metadados, conjuntos de dados compostos (*data sets*) e outras publicações ampliadas que guardam formas de relações entre si.

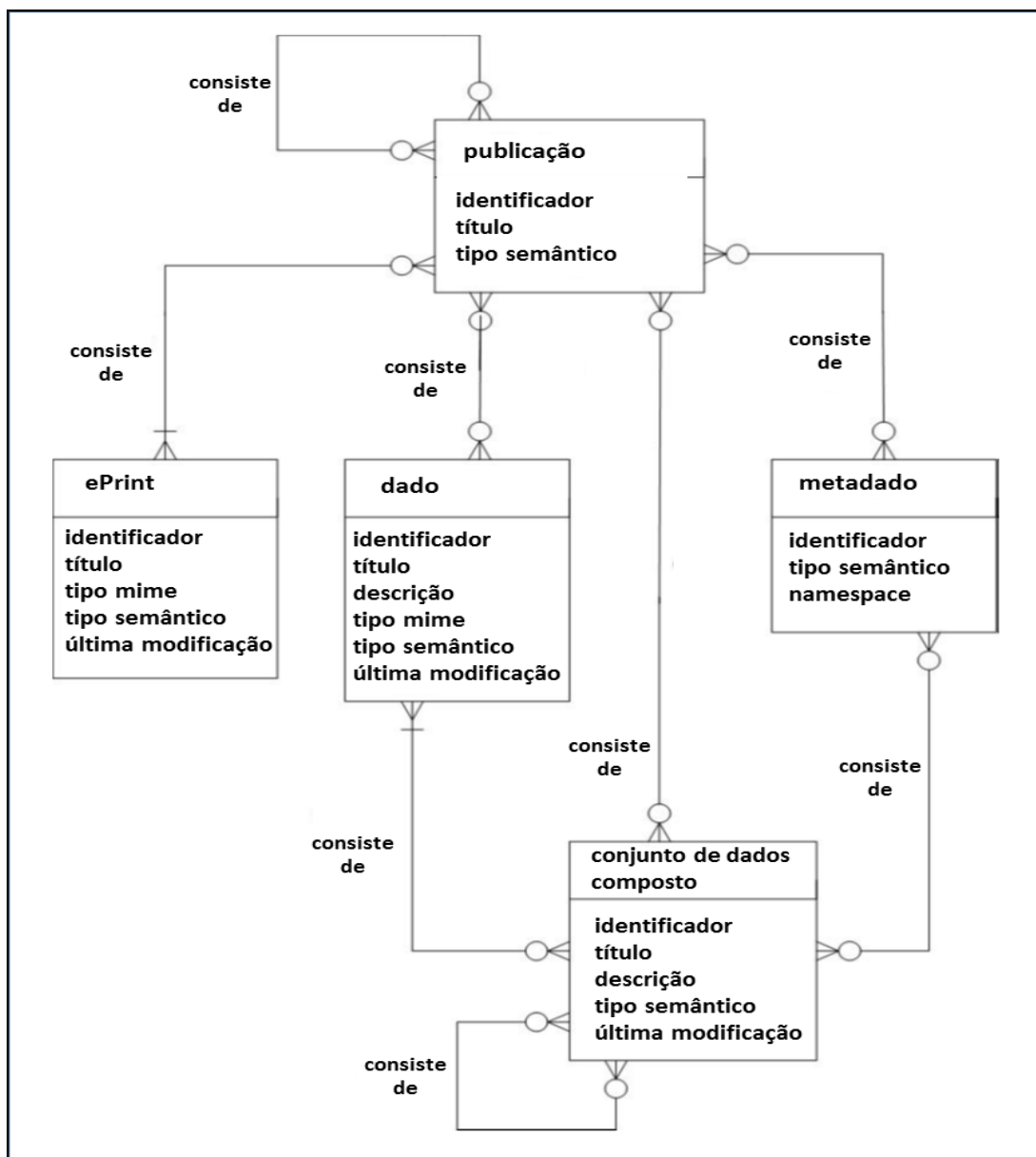


Figure 2: Diagrama entidade-relacionamento para entidades básicas e propriedades de publicações ampliadas.
Fonte: VEHAAR, 2007 (tradução nossa)

Muitas visões se sobrepõem em torno de uma publicação que é construída dinamicamente, e que traz subjacente a interlocução de muitos atores presentes e futuros. Esta característica determinante pode ser verificada em outro relatório do DRIVER-II, publicado em conjunto com o SURF-Foundation em 2009, intitulada “*Enhanced Publications: Linking Publications and Research Data in Digital Repositories*”, onde é descrita, entre outras questões, a estrutura de uma publicação ampliada formada pela agregação de elementos fornecidos pelo autor, pelos revisores e pelo editor da publicação. (VERNOOY-GERRITSEN, 2009)

De acordo com Lagoze e Van de Sompel (2007), os componentes estruturais de uma publicação fornecidos pelo autor são: tipo de (por exemplo, artigo, simulação, vídeo, conjunto de dados, software); tipo de mídia (por exemplo, texto, imagem, áudio, vídeo, misto); formato de mídia (por exemplo, PDF⁸, XML⁹, MP3¹⁰); endereço da rede, dado que componentes diferentes são acessados em diferentes repositórios.

De acordo com os autores acima citados, os elementos fornecidos pela revisão por pares no processo de ampliação da publicação tornam a revisão mais dinâmica, aberta e responsiva. Os pesquisadores crescentemente tem tido oportunidade de adicionar informações a artigos publicados, criando novas formas de socialização em torno de uma avaliação aberta. Podemos considerar os elementos: notas *on-line*, comentários, discussão interativa, *link* para blogs e grau conferido ao artigo, como pode ser observado na prática, no formato da revista Cell¹¹.

Ainda de acordo com com Lagoze e Van de Sompel (2007), os elementos fornecidos pelos editores compreendem todas as facilidades e ferramentas que eles possam disponibilizar para melhorar a precisão da recuperação da informação, aumentando a visibilidade do artigo e o interesse do leitor. O elemento considerado mais importante nesse fluxo é a marcação XML, e particularmente, o uso de linguagens de marcação orientadas para áreas específicas do conhecimento – por exemplo, a *Chemical Markup Language*. As linguagens de marcação facilitam a descoberta, indexação, disseminação e a vinculação por *links* para fontes externas. O uso de normas e padrões para publicação, de vocabulários controlados ou *link* para ontologias e ainda de interface ergonômicas permite o usuário identificar facilmente outros recursos que estão agregados.

O relatório do SURF-*Foundation* (VERNOOY-GERRITSEN, 2009)descreve algumas propostas de modelo para publicações ampliadas e enfatiza que em todos eles há uma tentativa de adicionar estrutura pelo uso de metadados nos mais diferentes níveis. Os objetos agregados destes modelos, dessa forma, se tornam integrados semânticamente, tornando, esta estrutura importante não somente para a leitura e compreensão humana, mas também para sua legibilidade pela máquina, permitindo que aplicações computacionais como, por exemplo, mineração de dados, possam explorar o potencial informacional das agregações.

⁸ PDF - *Portable Document Format*

⁹ XML- *eXtensible Markup Language*

¹⁰ Mp3 - *Moving Picture Experts Group*

¹¹ Disponível em: <<http://www.cell.com/>> Acesso em 27 jun. 2013.

O modelo apresentado por Kircz (2002) é uma forma modular de publicação: o artigo é dividido em módulos conectados entre si. Um módulo pode ser uma parte do artigo, por exemplo, *abstract*, informação bibliográfica ou pode ser um conjunto de dados ou de informações em outras mídias. Um módulo é definido como uma "representação única caracterizada independente de uma unidade de informação conceitual que visa comunicar essa informação" (HARMSZE, 2000 apud KIRCZ, 2002). Cada módulo possui seu conjunto de metadados, o que permite que cada parte da unidade de informação possa ser citada separadamente, reutilizada para diferentes propósitos ou ainda recombinação formando novos documentos complexos.

É esperado também que "imagens estáticas e em movimento, sons, simulações e em breve também informação tátil possam ser trocadas, experimentadas, e, portanto, analisadas e interpretadas, por diferentes pessoas em diferentes ambientes culturais e épocas" (KIRCZ, 2002). Todos estes componentes do documento eletrônico devem estar de acordo com normas que garantam sua qualidade e integridade.

Outra proposta de infraestrutura para publicações ampliadas é a de Seringhaus e Gerstein (2007), que sugerem que as publicações ampliadas devem:

- a) capturar uma gama de dados em formato digital e facilitar seu depósito em banco de dados juntamente com a publicação do manuscrito;
- b) indexar todos os artigos de periódicos em texto completo, associando palavras-chave e identificadores com os registros do banco de dados e *links* para livros, sítios de laboratório e comentário de alto nível;
- c) fornecer vários níveis de revisão por pares, comentários da comunidade e anotação;
- d) produzir artigos totalmente legíveis por máquina, fornecendo marcação inteligente e resumos digitais estruturados;
- e) ser uma rede de informações acessíveis através de um único portal.

No que tange às publicações ampliadas, os formatos de metadados consagrados, como Dublin Core e MARC não são adequados, pois não permitem a representação de agregações. Como estamos lidando com objetos, propriedades, tipos de objetos, relacionamento entre objetos, é conveniente uma estrutura de grafos ao invés de uma estrutura em árvore. A norma *Object Reuse and Exchange* definido pela *Open Archive Initiative* - conhecido pela sigla OAI-ORE - se torna adequado ao problema e será discutido a seguir.

3 A NORMA OAI-ORE

A *Web* é construída sobre a noção de unidades atômicas de informação, também chamada de "recursos", que são identificados por meio de URIs, que dão a elas um caráter único permitindo que ele seja encontrado de forma persistente. Além destas unidades atômicas, os recursos, muitas vezes são agregações de unidades atômicas formando um objeto complexo. O conjunto de especificações do OAI-ORE define padrões para a descrição e intercâmbio destas agregações de recursos *Web*. A página *Web* do OAI-ORE¹² explicita claramente a intenção da norma:

“Esta agregação, algumas vezes chamada de objetos digitais compostos, pode combinar recursos distribuídos com tipos múltiplos de mídia, incluindo texto, imagens, dado e vídeo. O objetivo da norma é expor o conteúdo rico dessa agregação para aplicações que suportem sistemas de autoria, depósito, intercâmbio, visualização, reuso e preservação”

A norma equaciona o problema crítico que é a ausência de forma padronizada para descrever os elementos constituintes do objeto digital composto e ainda os limites de uma agregação (LAGOZE; VAN DE SOMPEL, 2007).

A publicação ampliada é, na sua essência, um objeto complexo, composto pela agregação de diversos recursos digitais, que expõem uma relação entre si, e que podem estar distribuídos por diversos repositórios, comportando uma forte similaridade com o modelo abstrato de dados da norma OAI-ORE. Dessa forma, a aplicação da norma OAI-ORE, apesar de não ter sido criada para este fim, se torna conveniente.

As especificações do OAI-ORE são baseados no modelo *Object Reuse and Exchange* (ORE), que apresenta como característica determinante um Mapa de Recursos. Esse Mapa torna explícita a associação de uma identidade a uma agregação de recursos e faz declarações sobre a sua estrutura e semânticas. O formato de serialização primário para mapear o recurso é o perfil da unidade atômica expressa em *Resource Description Framework* (RDF) – uma linguagem para representação da informação na *Web* recomendada pelo consórcio W3C, para descrever por meio da sintaxe XML três elementos (recurso, propriedade e valor). Por ser expresso em RDF, o Mapa de Recursos também pode ser serializado por qualquer outro formato que permita serialização em RDF, o que permite uma aproximação com a Web Semântica.

¹² Disponível em: <www.openarchives.org/ore> Acesso em: 29 maio 2013.

Dessa forma, o Mapa de Recursos da agregação é uma forma de representar um conjunto de recursos associados por meio de relações. A figura 3 exhibe o modelo básico do OAI-ORE, que se resume em três entidades: a “Agregação”, que é a coleção de recursos web, que individualmente são chamados de “Recursos Agregados” e o “Mapa de Recursos” que é a entidade que contém a descrição de uma “Agregação”. Existem cinco propriedades que relacionam essas entidades: “descreve”, “é-descrito-por”, “agrega”, “é-agregado-por” e uma última que denota que dois recursos são idênticos.

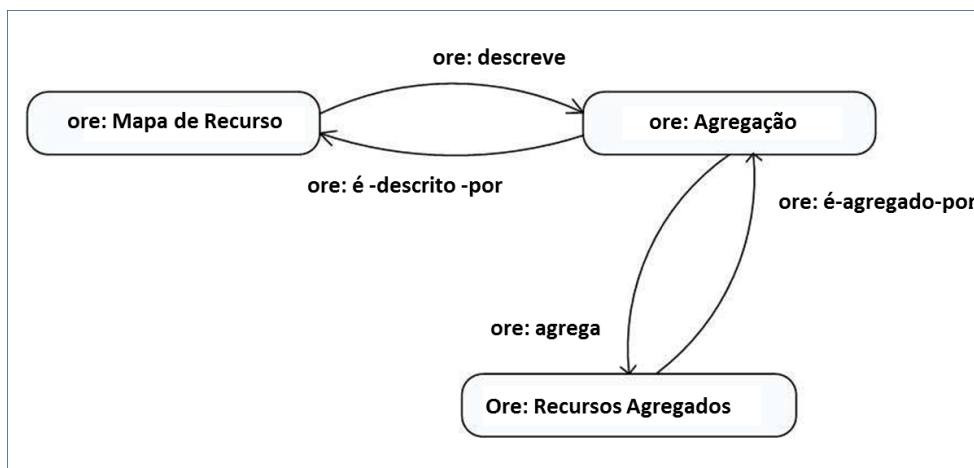


Figura3: Modelo Básico do OAI-ORE

Fonte: VERNOOY-GERRITSEN, 2009a (tradução nossa)

O Mapa de Recurso é o documento através do qual a publicação ampliada pode ser acessada, ele é usado, entre outras coisas, para “expor aos clientes de coleta automática (*harvesting*) os objetos compostos que ele oferece para acesso”, como explica Lagoze e Van de Sompel (2007). De uma forma direta, o Mapa de Recurso referencia toda a publicação ampliada.

A norma de empacotamento e intercâmbio de metadados *Metadata Encoding Transmission Standard* (METS) considera que os objetos complexos devem ser modelados como estruturas de árvore, assim como um livro com subcapítulos. Cada nó desta árvore deve ser associado a um conjunto de metadados expressos em XML. Este é o ponto de partida para o entendimento do OAI-ORE, ou seja, a compreensão de cada unidade de um objeto complexo deve ser descrita separadamente com metadados que registrem sua estrutura, linhagem e identificação persistente.

Diferente da outra norma do *Open Archive Initiative-Protocol for Metadata Harvesting*, a OAI-PMH, a OAI-ORE é considerada mais rica por requerer dos objetos digitais uma camada de interoperabilidade que vai além dos metadados puramente descritivos, considerando também na representação de cada elemento os metadados de preservação e administrativos. Uma vantagem importante do OAI-ORE é que ele pode ser adotado para encapsular recursos distribuídos. É interessante observar que o foco do OAI-ORE está mais no recurso e menos nos repositórios.

O conceito de Mapa de Recurso está ancorado em grafos RDF, que descrevem agregações e atributos para um formato legível por máquina. A figura 4 ilustra como uma instância de agregação – composta por uma publicação ampliada que agrega um *e-print*, dados e metadados – pode ser representada pelo modelo entidade-relacionamento, como pode ser descrita usando o vocabulário OAI-ORE e ainda como ela pode ser expressa em RDF/XML, aproximando-se do padrão tratado pela Web Semântica.

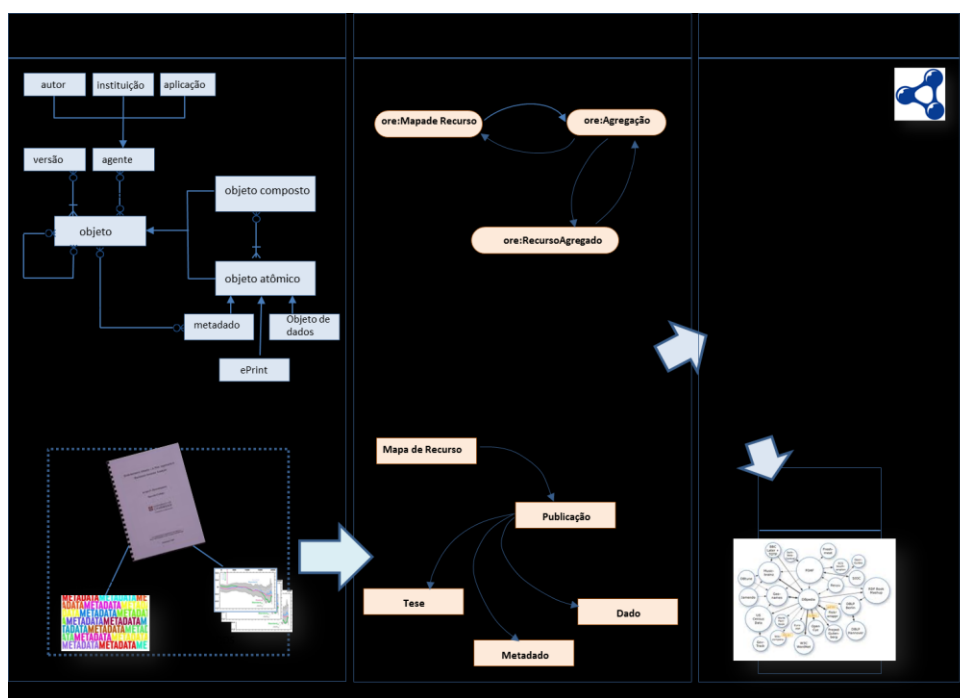


Figura 4: Publicação ampliada representada pelo modelo entidade-relacionamento, pelo modelo OAI-ORE e expresso em RDF/XML

Fonte: o autor, baseado em VERNOOY-GERRITSEN, 2009a

4 FINALIDADES E VANTAGENS

As publicações científicas e os dados de pesquisa são atualmente disseminados e compartilhados separadamente; publicar esses recursos de forma integrada torna-se uma opção cada vez mais real e traz inúmeras vantagens para o mundo da pesquisa científica. Entretanto, essa integração para ser efetiva, depende fortemente de aplicação de metadados.

Vejamos algumas vantagens dessa modalidade de publicação:

- a) a recuperação de *e-prints* e dados de forma integrada poupa o tempo gasto na busca por informações, pois se recupera um recurso e todos os recursos associados a ele;
- b) os dados são contextualizados pela própria publicação, mantendo seu sentido original, mas podendo ser reutilizados para novas pesquisas e reinterpretados em outros contextos;
- c) os dados passam a ser preservados para uso e reuso futuros, assim como as publicações tradicionais;
- d) os dados podem ser citados isoladamente e como um todo, de acordo com a finalidade da pesquisa;
- e) os dados permitem verificar a veracidade da pesquisa, apoiam os processos de replicação e dão mais confiabilidade aos seus resultados;
- f) as publicações ampliadas podem melhorar e tornar mais abertos e interativos os métodos de revisão por pares;
- g) a possibilidade de reuso dos dados financiados por recursos públicos abre possibilidades de encurtamento do ciclo de comunicação científica e de otimização dos recursos destinados à pesquisa.

Em suma, a perspectiva promissora de se ampliar as publicações científicas por meio de agregações bem descritas e com alto grau de formalização, de maneira que se tenham disponíveis dados de pesquisa associados com metadados confiáveis e consistentes, torna possível a efetiva recuperação, interpretação e reuso desses valiosos recursos informacionais.

5 À GUIA DE CONCLUSÃO

No ambiente de pesquisa baseado na geração e uso intensivo de dados, desencadeado principalmente pelo desenvolvimento de instrumentos científicos e pela aplicação de computadores e simulação, torna-se essencial a concepção de novos tipos de publicação que consigam expressar a sofisticação e a diversidade de tipos de conteúdo produzidos como resultado da pesquisa científica e acadêmica é uma demanda cada vez mais explícita dos pesquisadores. Nos últimos tempos - na busca de soluções para seus problemas de acesso e disseminação de resultados de pesquisa – os pesquisadores têm sido os principais agentes de inovação na área de informação. O protocolo OAI-PMH, o movimento do acesso livre e a tecnologia de repositórios digitais são exemplos para este fenômeno.

Os novos conceitos de publicação acadêmica trazem avanços significativos para a área de Ciência da Informação, pois colocam um problema contemporâneo de grande impacto e abrangência para a atividade de pesquisa, ao mesmo tempo em que reforçam os pressupostos da Ciência da Informação. Nessa perspectiva, a organização do conhecimento, a comunicação científica, a recuperação e a disseminação de informação se tornam áreas de forte relevância para a pesquisa científica como um todo.

O modelo aqui apresentado pode ser considerado uma novidade em termos de aplicação e estudos no âmbito da Ciência da Informação no país. Porém, neste trabalho, pretendeu-se apenas apresentar o conceito de “publicações ampliadas”, esclarecendo suas características, estrutura e finalidade.

Por fim, é importante ressaltar que estudos iniciais já apontam para alguns desafios que podem se inserir na agenda de pesquisa sobre esta temática, e que merecem ser investigados academicamente com mais profundidade e de forma interdisciplinar, pois o espectro é amplo, indo da esfera tecnológica até esfera política, social e ética, para citar algumas possíveis facetas. Alguns problemas em pauta:

- a) as relações entre coleções de dados e *e-prints* são muito elementares e precisam ser efetuadas de forma mais semântica;
- b) necessidade de desenvolvimento de mecanismos de recuperação, apresentação (renderização) e navegação dos dados e suas ligações, através de mecanismos visuais que usem taxonomia ou outras ferramentas de visualização semântica;
- c) tensão latente e não resolvida entre confidencialidade e acesso aos dados, colocando a questão de propriedade intelectual das agregações e de suas partes em v, além das questões éticas em voga.

- d) a ausência de flexibilidade ou extensibilidade, que permitam o cientista escolher interativamente um conjunto de recursos gerados a partir de uma experiência ou de investigação,
- e) apoio dos sistemas de publicação existentes somente quanto ao acesso livre de publicações e não dos dados;
- f) investigação de formatos de metadados para a gestão da preservação do significado e da estrutura de coleções de dados de pesquisa em ambientes de repositórios de dados;
- g) estudo de questões transversas a toda área, como interoperabilidade e integração com sistemas convencionais e sistemas avançados, como web semântica e *linked data*¹³;
- h) uso de ferramentas como ontologias que permitam não apenas a recuperação dos dados, mas o processamento de conhecimento de forma inteligente;
- i) impactos no ciclo de comunicação científica, compartilhamento de recursos e trabalhos cooperativos, entre outros.

A identificação das problemáticas de pesquisa acima mencionadas é o principal resultado dos estudos realizados até agora e se configura em uma agenda de pesquisa a ser perseguida na busca de soluções para os novos desafios emergidos a partir do quarto paradigma e também dessa nova ferramenta de comunicação científica.

REFERÊNCIAS

- AALBERSBERG, Ijsbrand Jan; DUNHAN, Judson; KOERS, Hylke. *Connecting scientific articles with research data: new directions in online scholarly publishing*. 2011. Disponível em: <<http://www2.nict.go.jp/isd/ISDS-contents/wds-kyoto-2011.org/pdf/IS704.pdf>> Acesso em: 19 maio 2013.
- CHEUNG, Kwok et al. SCOPE: a scientific compound object publishing and editing system. *International Journal of Digital Curation*, v. 3, n. 2, p. 4-18, 2008.
- ERIKSSON, Henrik. The semantic-document approach to combining documents and ontologies. *International Journal of Human-Computer Studies*, v. 65, n. 7, p. 624-639, 2007.
- KIRCZ, Joost G. New practices for electronic publishing 1: Will the scientific paper keep its form? *Learned Publishing*, v. 14, n. 4, p. 265-272, 2001. Disponível em: <www.kra.nl/Website/Artikelen/learned-publ1.pdf> Acesso em: 11 ago. 2013

¹³ Linked data é uma iniciativa do W3C para formar um conjunto de práticas para publicar e interligar dados estruturados na Web.

KIRCZ, Joost G. New practices for electronic publishing 2: New forms of the scientific paper. *Learned Publishing*, v. 15, n. 1, p. 27-32, 2002. Disponível em: <www.kra.nl/Website/Artikelen/Learnedpubl2.pdf> Acesso em: 11 ago. 2013

LAGOZE, Carl; VAN DE SOMPEL, Herbert. Compound Information Objects: The OAI-ORE Perspective. *Open Archives Initiative*. 2007. Disponível em: <<http://www.openarchives.org/ore/documents/CompoundObjects-200705.html>>. Acesso em: 10 ago. 2013.

SAYAO, L. F. ; SALES, L. F.. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. *Informação & Sociedade (UFPB. Online)*, v. 22, p. 1, 2012. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/12224/8586>> Acesso em: 11 ago. 2013

SERINGHAUS, Michael R.; GERSTEIN, Mark B. Publishing perishing? Towards tomorrow's information architecture. *Bmc Bioinformatics*, v. 8, n. 1, p. 17, 2007. Disponível em: <www.biomedcentral.com/1471-2105/8/17> Acesso em: 11 ago. 2013

VAN DEN HEUVEL, Charles; RAYWARD, W. Boyd. Facing interfaces: Paul Otlet's visualizations of data integration. *Journal of the American Society for Information Science and Technology*, v. 62, n. 12, p. 2313-2326, 2011. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.21607/abstract;jsessionid=B9BB9F523488602613B93F6D07543A08.d03t02?deniedAccessCustomisedMessage=&userIsAuthenticated=false>> Acesso em: 11 ago. 2013

VAN DE SOMPEL, Herbert et al. An interoperable fabric for scholarly value chains. *D-Lib Magazine*, v. 12, n. 10, p. 1082-9873, 2006.

VERHAAR, Peter. *Report on object models and functionalities*. DRIVER II, 2007. Disponível em: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report_on_Object_Models_and_Functionalities.pdf?sequence=2> Acesso em: 20 dez. 2011.

VERNOOY-GERRITSEN, Marjan (ed.). *Enhanced publications: linking publications and research data in digital repositories*. Amsterdam University Press, 2009. Driver and Surf foundation studies. Disponível em: <dare.uva.nl/document/150723> Acesso em: 11 ago.

VERNOOY-GERRITSEN, Marjan (ed.). *Emerging standards for enhanced publications and repository technology: survey on technology*. Amsterdam University Press, 2009a. Driver and Surf studies. Disponível em: <dare.uva.nl/aup/en/record/316870> Acesso em: 11 ago. 2013

WOUTERSEN-WINDHOUWER, Saskia; BRANDSMA, Renze et al. *Enhanced publications: state of the art*. Part 1, p.19-91, 2009. Disponível em: <www.driver-repository.eu/Enhanced-Publications.html> Acesso em: 11 ago. 2013

WULF, William A. The national collaboratory—a white paper. *Towards a national collaboratory*, p. 17-18, 1989.