

NANOPUBLICAÇÕES: ABORDAGENS DE REPRESENTAÇÃO DO CONHECIMENTO COMO INOVAÇÃO PARA A COMUNICAÇÃO CIENTÍFICA

NANOPUBLICATIONS: KNOWLEDGE REPRESENTATION APPROACHES AS AN INNOVATION FOR SCIENTIFIC COMMUNICATION

Luis Fernando Sayão^a
Luana Farias Sales^b

RESUMO

Objetivo: Este estudo tem como objetivo apresentar a noção de nanopublicação, um constructo apresentado pela *Concept Web Alliance*, cujo conceito está baseado na divisão do conhecimento em afirmativas com proveniência e contexto, tornando-se uma unidade mínima publicável e citável, estruturado em RDF e em outras tecnologias semânticas que permitem a interpretação por humanos e a acionalidade por máquina.

Metodologia: o estudo, de caráter exploratório e bibliográfico, está baseado em autores que implementaram os fundamentos teóricos e práticos da proposta. **Resultado:** a análise realizada permite concluir que a criação de servidores e redes de nanopublicações em conjunto com conceitos mais recentes, como FDO e IFDS, tem o potencial de mudar o cenário da comunicação científica, se constituindo menos numa ruptura do modelo clássico e mais numa expansão técnica e conceitual necessária às dinâmicas da *eScience*. **Conclusão:** Conclui-se que a adoção de nanopublicações e conceitos como FDO e IFDS oferece uma evolução essencial para a comunicação científica na *eScience*, promovendo uma estrutura mais eficiente e interoperável entre humanos e máquinas.

Descritores: Nanopublicação. Tecnologias semânticas. Comunicação científica.

^a Doutor em Ciência da Informação pela Universidade Federal do Rio de Janeiro (UFRJ/IBICT). Comissão Nacional de Energia Nuclear (CNEN) – PPGCI IBICT – UFRJ; PPGB – UNIRIO. E-mail: sayao@cnen.gov.br

^b Doutora em Ciência da Informação pela Universidade Federal do Rio de Janeiro (UFRJ/IBICT). Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), PPGCI IBICT – UFRJ; PPGB – UNIRIO. Email: luanafsales@gmail.com

1 INTRODUÇÃO

Nas últimas décadas, o *corpus* de conhecimento acadêmico coletivo - tanto textos narrativos como dados publicados e outros produtos de processos de pesquisa - excedeu a compreensão humana, desafiando a capacidade dos pesquisadores, mesmo daqueles que trabalham em domínios disciplinares estritamente definidos, de se manterem atualizados e de integrar os novos avanços aos seus espaços epistemológicos (Schultes *et al.*, 2022).

No âmbito da ciência intensiva em dados, o abismo entre a coleta/geração e as etapas de processamento e análise desses dados se torna cada vez maior. Nessa condição, muitos dados passam despercebidos ou se tornam opacos para o escrutínio, interpretação e reanálise pela pesquisa e por outros segmentos sociais. Eles falham em revelar as informações neles contidas ou permanecem inacessíveis devido a vários fatores, tornando uma parcela cada vez maior da ciência autocontida desconectada e invisível para o consumo de seres humanos e para interpretação e ações de provedores de serviços computacionais autônomos. Como resultado, a inconsistência, a ambiguidade e a redundância de dados e informações na *web* estão se tornando impedimentos para o desempenho da extração e análise abrangentes de informações (Mons; Velterop, 2009).

Nesse cenário, a busca por uma informação específica se torna difícil pela maneira como as publicações são arquivadas, comumente em formato textual não estruturado e sem o uso de técnicas de marcação e de destaque de seus conteúdos. Dessa forma, as informações se tornam ilegíveis e não acionáveis por computadores, dificultando a pesquisa, preservação, agregação de valor e o reúso, e impedindo a máxima exploração do seu potencial informacional (Mons *et al.*, 2011; Silva; Bordin, 2020).

Esse alto grau de entropia informacional que se instala nos processos científicos contemporâneos tem um rebatimento nítido nos padrões de comunicação científica, na medida em que a quantidade de comunicações acadêmicas aumenta vertiginosamente tornando cada vez mais difícil que afirmações científicas essenciais e específicas sejam encontradas, conectadas

e curadas. Além disso, a redundância dessas declarações em vários fóruns dificulta determinar atribuição de autoria, qualidade e procedência de suas declarações. Isto se deve, pelo menos em parte, ao fato de que estamos lidando ainda com um paradigma arcaico de publicações acadêmicas, que são, de forma otimista, fáceis de serem lidas por seres humanos especialistas, mas que permanecem, na maior parte dos casos, inacessíveis para procedimentos automatizados. Este fenômeno do nosso tempo fica mais claro quando nos referimos ao espaço de significação da *eScience*, que impõe a condição determinante de que somente um alto grau de automação pode fazer frente ao volume e velocidade dos seus fluxos de geração de dados e informações (Wilkinson *et al.*, 2016).

Atualmente, os artigos são rotineiramente disponibilizados em repositórios digitais, mas ainda assim o conteúdo e a forma permanecem vinculados aos leitores humanos, enquanto as poderosas capacidades de processamento de informações das máquinas, que deveriam ajudar o pesquisador, são direcionadas à cálculos circunstancialmente irrelevantes, como pontuações de impacto (Schultes, 2024). A publicação acadêmica secular, em papel, mesmo quando convertida para formatos eletrônicos, nem sequer começou a explorar seriamente as possibilidades que a *Web Publishing* oferece, mesmo em seu estágio atual, ainda em desenvolvimento. Assim sendo, a maioria das chamadas publicações eletrônicas disponíveis são meros simulacros das versões em papel, e frequentemente se apresentam apenas em formato PDF (Mons; Velterop, 2009) constituindo, num exagero estilístico, num prosaico “mimeógrafo digital”.

Para enfrentar esses desafios, a *Concept Web Alliance* promoveu a noção de nanopublicação, na qualidade de declarações científicas essenciais com proveniência e contexto associados que representam unidades mínimas de dados estruturados e constitui uma entidade citável, cuja arquitetura é implementada por meio de tecnologias semânticas (Groth; Gibson; Velterop, 2010; Mons; Velterope, 2009), e são legíveis e acionáveis por máquina.

O presente ensaio tem como objetivo apresentar o conceito de nanopublicação contextualizando-o no domínio da comunicação científica contemporânea, que se realiza como uma forma mais eficiente de comunicação

de conhecimento nos ambientes intensivos em dados, sustentados por infraestruturas computacionais e de rede. O estudo tem um caráter exploratório e bibliográfico, e está ancorado principalmente nas publicações de autores e instituições que desenvolveram os fundamentos teóricos e práticos da nanopublicação e do seu entorno tecnológico.

2 O PROBLEMA: O CONTEXTO DE EXCESSO DE PUBLICAÇÕES ACADÊMICAS

O vertiginoso avanço das tecnologias computacionais reconfigura de forma pervasiva os fluxos de praticamente todas as atividades sociais públicas e privadas, como governo, negócios, indústria, arte, lazer, e as interações entre as pessoas e organizações. A ciência, pelo seu envolvimento natural com as tecnologias, cujas bases estão apoiadas nas inovações proporcionadas pelo progresso científico, é intensamente afetada na sua capacidade de observar as dinâmicas da natureza e da sociedade, de criar simulações e virtualizações multifacetadas do universo e de construir e reconstruir novos métodos de análise. Esse deslocamento para o abstrato engendra uma ciência inteiramente nova, plataformizada por ciberinfraestruturas de pesquisa, que pode ser resumida pelo termo *eScience* (Gray *et al.*, 2005; Hey; Tansley; Tolle, 2009), e caracterizado como “um meio que permite acesso e circulação de conhecimento distribuído, em que colaboram e se comunicam diferentes comunidades e disciplinas, rompendo fronteiras culturais, geográficas e temporais”, como esclarece Pérez-González (2010, p. 3), é uma nova forma de cultura científica que se sustenta em uma robusta infraestrutura tecnológica de alto nível, cujo deslocamento simbólico pode ser caracterizado pela hipérbole que a cada dia que passa se torna mais verdadeira: passamos de descobertas *in vitro* e *in vivo* para descoberta *in silico*.

Deve ficar claro que essas rupturas não se limitam somente às infraestruturas de pesquisa e à conceituação metodológica, para trabalhar nesses novos espaços de significação, uma nova geração de e-cientistas está surgindo. Eles estão criando formas inéditas de trabalhar, entendem profundamente as possibilidades das tecnologias e realizam suas pesquisas,

não como um ser humano individual, mas como um nó em uma rede de humanos e máquinas. Este fenômeno começa a colocar em xeque os padrões tradicionais da comunicação científica baseados em periódicos científicos (Sales; Sayão, 2023; Wouters, 2006).

Paradoxalmente, essa apropriação das tecnologias digitais pela pesquisa científica não encontra o mesmo nível de assimilação pelos fluxos de comunicação científica, mesmo diante do grande volume de produtos de pesquisa que precisam ser disseminados, conectados e reusados para integralizar as etapas do ciclo de vida da pesquisa e acelerar o seu progresso, mesmo diante das inovações tecnológicas dos meios de comunicação e mesmo diante das mudanças sociológicas da ciência.

Por séculos, a publicação científica é o meio pelo qual as comunidades científicas comunicam, compartilham e avaliam o conhecimento científico nos seus espaços de significação e, idealmente, traduzem os enigmas e os achados da ciência para toda a sociedade. Quando observamos o panorama das publicações acadêmicas clássicas, fica evidente que elas completaram o processo de digitalização e estão sob um novo regime universal de exploração econômica e consumo. Todavia, essa transformação do impresso para o eletrônico/digital não oferece uma contrapartida por meio da construção de concepções que respondam as demandas da pesquisa contemporânea, intensiva em dados e performada *in silico*, e um alinhamento ao progresso tecnológico, mais especificamente aos avanços das tecnologias semânticas. “Mesmo quando convertida para formatos eletrônicos, não se começou a explorar seriamente as possibilidades abertas pela publicação na *web*, mesmo em seu estágio atual, que se encontra ainda em desenvolvimento” (Mons; Velterop, 2009). A maioria das chamadas publicações eletrônicas disponíveis são meros análogos das versões em papel e, muitas vezes, estão apenas formatadas em PDF, voltadas ciclicamente para o retorno à forma impressa.

Apesar do fato incontestável de que a tecnologia mudou a forma e a velocidade de como a ciência realiza e dissemina suas pesquisas, há um descompasso entre a ciência digitalizada e as suas formas assumidas de comunicação e de avaliação - elos imprescindíveis no ciclo de formalização do

conhecimento - que reiteradamente clamam por práticas mais eficientes, transparentes e colaborativas que melhorem os paradigmas centenários de periódicos, artigos e de revisões tradicionais por pares, especialmente na conjuntura de uma ciência intensiva em dados e representações, onde o “texto não é o único nem o mais efetivo meio de compartilhar informações” (Mons *et al.*, 2011, p. 281).

O que se observa é que esse elo imprescindível do ciclo da comunicação científica atualmente carece de representações interpretáveis por máquina. Os artigos científicos são publicados em textos longos, com uma alta granularidade – que define um baixo nível de detalhamento das assertivas – apresentando estruturas complexas que são presas a formatos otimizados para interpretação e consumo de leitores humanos e não para meios automatizados de organização e acesso, observam Cristina-Iulia Bucur, Kuhn e Ceolin (2020). Além do mais, não estamos usando os avanços já consolidados que estão disponíveis, como as tecnologias em torno da *web* semântica e dados ligados. O alvorecer da *web* semântica trouxe a primeira onda de redução de ambiguidades na estrutura da *web*, na medida em que conectam os seus recursos no nível do 'conceito', em vez do nível do termo, como feito pelo Google e por outros sistemas baseados em palavras, este parece ser um ponto de partida importante para novas formulações de comunicação científica e descoberta de recursos (Mons; Velterop, 2009).

Não se argumenta aqui que as formas narrativas da comunicação acadêmica devam ser abolidas, posto que elas são necessárias tanto agora quanto no futuro, como foram imprescindíveis também no passado, mas precisamos também reconhecer que a ciência intensiva de dados precisa de informações acionáveis por máquina e interpretáveis por *stakeholders* computacionais que atuem e prestem serviços de integração e análises em nome dos pesquisadores, frente à incapacidade humana de processar o acúmulo de dados e informações na velocidade necessária ao avanço da ciência, e que novas formulações de comunicação possam definir um ecossistema informacional mais coerente com as necessidades da pesquisa e com os avanços tecnológicos.

À medida que a quantidade de comunicações acadêmicas aumenta, é cada vez mais difícil que declarações científicas fundamentais e específicas sejam encontradas, acessadas e integradas aos seus domínios epistemológicos, e seus requisitos de qualidade e proveniência sejam formalmente declarados. A busca por uma informação específica se torna difícil pela maneira como as publicações são armazenadas, comumente em formato textual não estruturado, e sem o uso de técnicas de marcação e de destaque de seus conteúdos. Dessa forma, as informações se tornam ilegíveis para computadores, dificultando a pesquisa, preservação, agregação de valor e o reuso, dificultando a extração máxima do seu potencial informativo (Mons *et al.*, 2011; Silva; Bordin, 2020).

Diante desse quadro podemos alinhar os seguintes problemas com as publicações tradicionais, no domínio de pesquisas orientadas por dados.

- A inconsistência, a ambiguidade e a redundância de dados e informações na *web* que se tornam obstáculos para o desempenho da extração e análise abrangentes de informações (Mons; Velterop, 2009);
- O volume crescente de comunicações torna difícil a recuperação de assertivas científicas específicas;
- Textos formatados para a lógica da leitura por humanos e para impressão (como o PDF);
- Apresentação desestruturada das publicações tradicionais que dificulta a interpretação e a acionabilidade por *stakeholders* computacionais;
- Pouca possibilidade de aplicação de tecnologias semânticas às comunicações científicas;
- Desconexão de anotações, comentários e revisões assinaladas à comunicação e desagregação em relação a outras comunicações relacionadas ao mesmo tema.

Com o propósito de equacionar soluções abrangentes para esses desafios, por meio de processos de publicação científica, com um nível de resolução granular mais refinado, que tornem mais visíveis e conectadas as

suas assertivas fundamentais, utilizando para tal, registro semântico formal de cada elemento, de maneira que seja possível alinhar com mais eficiência as publicações científicas com os princípios da *web* semântica e do *linked data*, é proposto pela *Web Alliance Concept*, uma abordagem que possibilite a concretização do conceito de nanopublicação.

Trata-se de um modelo unificador para representar de uma forma semântica os elementos das publicações, suas avaliações, bem como os processos envolvidos, atores e proveniência em geral. Apesar das conexões recentes das nanopublicações com os conceitos ainda em evolução do *FAIR Digital Object (FDO)* e da *Internet of FAIR Data and Service (IFDS)* (Santos *et al.*, 2023; Sayão; Sales; Carvalho Segundo, 2024), a ideia, como ressaltam as autoras Paula e Moura (2015), foi observada já em 2009, em Nova York, num evento que reunia especialistas em dados e tecnologias de *web* semântica.

A translação subjacente ao conceito de nanopublicação está em trocar o tratamento de grandes volumes de textos típicos das publicações acadêmicas, por representá-los na forma de pequenos trechos - por exemplo, parágrafos - que têm semântica formal incorporada e podem ser tratados como unidades de publicação independentes. Eles podem se linkar a outras unidades e, portanto, formar uma entidade maior - como um artigo completo ou revisão - formando uma rede complexa de *links* (Mons *et al.*, 2011, p. 261). Esses pequenos trechos de texto podem ser representados como nós em uma rede e podem ser vinculados uns aos outros formando assim redes de contribuições distribuídas e semanticamente anotadas, e estarem agrupados em servidores específicos.

Essas abordagens funcionam principalmente em artigos já publicados, mas tem sido argumentado que as descobertas científicas e seus contextos devem ser expressos em representações semânticas desde o início pelos próprios pesquisadores, no que foi chamado de publicação semântica genuína (Kuhn; Dumontier, 2017). Assim, novas modalidades de aplicações foram possíveis, como veremos posteriormente.

3 AFINAL, O QUE SÃO NANOPUBLICAÇÕES?

A nanopublicação estabelece um novo conceito de representar dados, informações e, em particular, afirmações científicas. Ela parte da ideia primordial de subdividir os resultados da pesquisa em pequenas partes denominadas de afirmações/assertivas, cujas representações devem estar numa notação formal baseada em RDF (Resource Description Framework) e portem informações de proveniência neste nível atômico, e que cada uma dessas pequenas entidades seja tratada como uma publicação separada, constituindo a menor unidade de informação publicável e citável (Kuhn *et al.* 2013). Por conseguinte, uma nanopublicação é uma afirmação cientificamente significativa sobre qualquer coisa que pode ser identificada e atribuída exclusivamente aos seus autores e serve para comunicar uma única declaração, sua fonte original (proveniência) e o registro de citação (informações da publicação) (Kuhn *et al.*, 2013; Mons, Velterop, 2009).

-Conceitualmente, as nanopublicações complementam as narrativas de conhecimento científico registradas em formatos textuais por pesquisadores, porém, diferente de um artigo de periódico convencional, ela é um pequeno trecho preciso e estruturado que expressa uma afirmativa científica e que existe como parte reutilizável e citável de um grafo de conhecimento crescente, cada afirmativa associa dois conceitos por meio de um predicado em formato digital - por exemplo, o medicamento X faz remissão da doença Y – acompanhados de metadados apropriados sobre procedência e contexto. Cada conceito em uma nanopublicação tem uma identificação única, universal e inequívoca, não semântica, estável, para a qual diferentes *Uniform Resource Identifiers* (URIs) podem ser resolvidos. As nanopublicações são publicadas em uma rede descentralizada de servidores, em um formato legível por humanos, mas também "compreensível" e acionável por computadores e seus algoritmos. Esta rede é composta por múltiplos servidores de nanopublicações implantados em várias localidades, como universidades e outras instituições de pesquisa (Nanopublications, 2022).

Estruturalmente, cada nanopublicação tem uma arquitetura que contém:

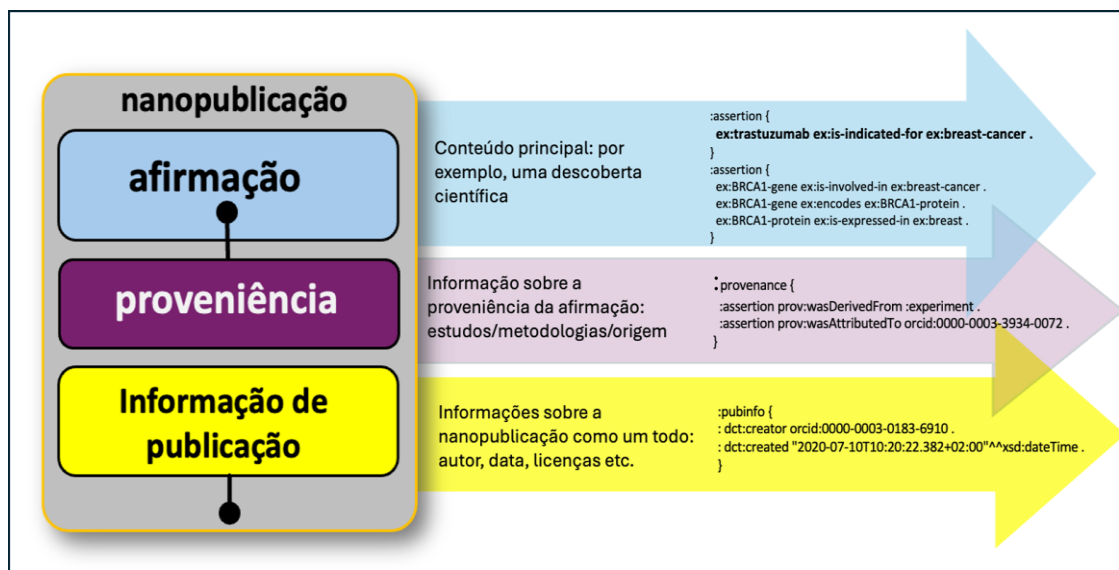
- **Afirmção:** é essencialmente uma declaração portando o conteúdo da nanopublicação – por exemplo, uma descoberta científica do tipo: *MALARIA é transmitida* por MOSQUITO (Groth; Gibson; Velterop, 2010). O grafo da afirmação da nanopublicação contém as afirmações que formam a afirmação principal.

- **Proveniência:** É uma indicação do estudo/metodologia/origem da afirmação. É constituída, por exemplo, por informações sobre que estudos foram conduzidos para derivar a afirmação, ou, de que documento ela foi extraída. O grafo de proveniência de uma nanopublicação contém uma ou mais triplas RDF que fornecem informações sobre a **afirmação**.

- **Informações sobre a publicação:** O grafo de informação sobre a publicação contém uma ou mais triplas RDF que fornecem informações de proveniência sobre a nanopublicação propriamente dita. São informações sobre a nanopublicação como um todo – por exemplo, autores, carimbos de tempo, URI, licenças, instituições envolvidas, que são essenciais para determinar a autoria por outros pesquisadores.

Cada uma das três partes de uma nanopublicação é representada como um grafo RDF, ou seja, uma coleção de triplas RDF, portanto, interpretáveis por máquina (Bucur; Kuhn; Ceolin, 2020) e expressas também em formatos compreendidos por leitores humanos, dessa forma constituem um conceito e uma tecnologia baseados em *linked data* e podem ser citados por meio de seus identificadores URI exclusivos. A figura 1 esquematiza essa estrutura e exhibe exemplo de codificação RDF.

Figura 1 – Arquitetura da nanopublicação



Fonte: os próprios autores baseados em Nanopublicatios (2022).

3.1 CARACTERÍSTICA E VANTAGENS

As nanopublicações podem ser usadas para expor dados quantitativos e qualitativos, bem como hipóteses, alegações, resultados negativos e opiniões que geralmente não são publicadas. Podem também disseminar dados individuais, como publicações independentes ou acompanhadas de artigo de pesquisa. Os diferentes níveis de proveniência permitem que os usuários avaliem a confiabilidade dos dados e fornecem um mecanismo pelo qual autores e instituições podem ser reconhecidos por sua contribuição ao mapa de conhecimento global. Além disso, como as nanopublicações podem ser atribuídas e citadas, elas fornecem incentivos para que os pesquisadores disponibilizem seus dados em formatos-padrão que impulsionam a acessibilidade e a interoperabilidade dos dados (Gray *et al.*, 2024).

Na perspectiva dos processos de apuração científica, como a revisão por pares, a segmentação em pequenos trechos permite levar em consideração a especialização dos revisores e o contexto detalhado de seus comentários de revisão. Esses comentários de revisão podem ser vinculados formal e precisamente às partes do artigo que abordam. Cada artigo, cada parágrafo e

cada comentário de revisão, portanto, forma um único nó em uma rede e é identificado por um URI (Bucur; Kuhn; Ceolin, 2020).

Ao facilitar o rastreamento de descobertas individuais até sua origem e/ou atualizações posteriores, as nanopublicações também ajudam a entender melhor a procedência dos dados científicos. Assim sendo, é possível garantir que a proveniência de cada segmento de informação possa ser rastreada com precisão junto com seu tempo de criação e autor e, portanto, permitir uma publicação mais flexível e eficiente do que o paradigma atual (Pensoft.Blog, 2023).

O formato e o fluxo de trabalho de nanopublicação garantem aos autores que as principais declarações científicas – aquelas que sustentam seu trabalho de pesquisa – sejam comunicadas de forma eficiente, tanto de forma legível por humanos quanto acionável por máquinas, de acordo com os princípios FAIR (Wilkinson *et al.*, 2016). Assim, suas contribuições para a ciência são mais bem preparadas para uma realidade impulsionada pelas tecnologias emergentes, especialmente a de inteligência artificial e algoritmos. A rede de nanopublicações interligadas também pode estabelecer um fórum valioso para cientistas testarem, compararem, complementarem e desenvolverem os resultados e abordagens uns dos outros, para um problema científico comum, ao mesmo tempo em que mantem o registro de sua cooperação em cada etapa do caminho. Além do mais, as nanopublicações são projetadas também para abordar a questão da avaliação da pesquisa num contexto em que as novas dinâmicas sociológicas da ciência ampliam o escopo da disseminação dos resultados iniciais de pesquisa, incluindo *preprints*, manuscritos aceitos e artigos não convencionais (Pensoft.Blog, 2023).

Numa outra dimensão, as nanopublicações foram projetadas para serem extensíveis, podendo ser modificadas conforme as necessidades de diferentes domínios e comunidades (Groth; Gibson; Velterop, 2010).

Em síntese, Cristina-Iulia Bucur, Kuhn e Ceolin (2020) destacam que um modelo de publicação unificado e semântico baseado em nanopublicações pode tornar a comunicação científica mais eficaz e amigável ao usuário. Além do mais, refirmam os autores: “Nossos resultados mostram que podemos representar

praticamente os diferentes elementos da comunicação científica, como artigos e revisões, com uma granularidade fina e de um modo semântico por meio das nanopublicações” (Cristina-Iulia Bucur, Kuhn e Ceolin, 2020, p.13)

4 TECNOLOGIAS E ARQUITETURA DAS NANOPUBLICATIONS

A nanopublicação é uma abordagem orientada pela comunidade científica para representar dados estruturados cuja acionalidade por máquina se caracteriza como um padrão para a sua operacionalização e consecução de seus objetivos. Na anatomia da nanopublicação, cada afirmação compreende um sujeito, um objeto e um predicado, entendido como um tipo de relação entre o sujeito e o objeto, que pode ser expressa por uma notação formal baseada em RDF, complementada por informações de proveniência e autoria da publicação. Além da RDF, a implementação da arquitetura das nanopublicações é realizada, por meio de outras tecnologias semânticas, tais como OWL e SPARQL.

Paul Groth, Gibson e Velterop, (2010) argumentam que, independentemente do domínio no qual a nanopublicação será aplicada, é necessário que existam boas descrições/representações do conhecimento que viabilize uma compreensão plena, uma recuperação eficaz e a integração e o reúso das informações. Por isso, a criação das nanopublicações deve ser guiada por **ontologias** que descrevam e estabeleçam relações entre os elementos e conceitos de um determinado domínio. Considera-se também que cada nanopublicação é única e tem sua própria identificação persistente.

Uma característica essencial das nanopublicações é que cada elemento referencie recursos externos padronizados e estruturantes. Assim sendo, as afirmações semanticamente formalizadas são expressas em termos consensados pela comunidade, por meio de *links* para recursos disponíveis *online*, que podem estar disponíveis abertamente como vocabulários controlados, ontologias e padrões

A acionalidade por máquina é essencial para a consecução do conceito de nanopublicação. O formato estruturado semanticamente das nanopublicações abre a possibilidade de que *stakeholders* computacionais – como algoritmos e assistentes com tecnologias IA - acessem os fragmentos de

conhecimento que elas pretendem comunicar, de forma que eles sejam totalmente compreensíveis e acionáveis. Com as nanopublicações, cada um desses fragmentos de informação científica é interconectado e rastreável até seu(s) autor(es) e as evidências científicas que os autenticam (Pensoft.Blog, 2023).

5 APLICAÇÕES

A arquitetura das nanopublicações e seus compromissos semânticos permitem diferentes configurações em torno do compartilhamento de afirmações para o consumo de agentes humanos e computacionais dentro de infossistemas semanticamente bem definidos. O periódico Biodiversity Data Journal (BDJ), por exemplo, explora essas possibilidades, conjugadas com a publicação acadêmica tradicional, de diversas formas, usando os fluxos de trabalho e modelos de nanopublicação disponíveis pelo periódico. Dessa forma, os autores do periódico podem fazer uso de três tipos de nanopublicações:

- **Nanopublicações associadas a um manuscrito submetido ao BDJ.** Este fluxo de trabalho permite que os autores adicionem uma seção de nanopublicações dentro de seu manuscrito enquanto preparam sua submissão. Basicamente, os autores ‘destacam’ e ‘exportam’ pontos-chave de seus artigos como nanopublicações para garantir ainda mais a FAIRificação das descobertas mais importantes de suas publicações.
- **Nanopublicação individual relacionada a qualquer publicação científica,** independentemente de seu autor ou fonte. Isso pode ser feito por meio da página “Nanopublications” acessível no site do BDJ. A principal vantagem da nanopublicação individual é que afirmativas científicas diretas se tornam disponíveis e FAIR antecipadamente, e permanecem prontas para serem adicionadas a um futuro artigo acadêmico.
- **Nanopublicações como anotações para publicações científicas existentes.** Ao anexar uma anotação ao artigo inteiro ou a uma seleção de texto, um leitor pode avaliar e registrar uma opinião sobre qualquer artigo usando um modelo simples baseado na *Citation Typing Ontology* (CiTO).

O exemplo a seguir – representado na figura 2 – ilustra a implementação de nanopublicação para a expansão do nível de interoperabilidade aplicada ao

projeto *Virus Outbreak Data Analysis Network – Africa* (VODAN–A) e ambientada dentro de uma internet de dados e serviços FAIR (Ferreira; van Reisen, 2023).

Figura 2 – Exemplo de nanopublicação aplicado ao projeto VODAN-A

The diagram shows a JSON-LD nanopublication structure divided into three sections:

- Assertion:** A blue box containing a JSON-LD snippet for `sub:assertion` and a plain language explanation: "The assertion: The product protein from gene ENSG00000103197 interacts through a particular pathway with a product of gene ENSG00000117020."
- Provenance:** An orange box containing a JSON-LD snippet for `sub:provenance` and a plain language explanation: "The provenance is a study on thyroid hormone production and their peripheral downstream signaling effects regarding congenital hypothyroidism."
- Publication Info:** A yellow box containing a JSON-LD snippet for `sub:publicationInfo` and a plain language explanation: "The nanopublication was minted in 2020 as a part of a Maastricht University study under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication."

On the right side of the diagram, there is a `Head` section with a JSON-LD snippet for `sub:head` and a plain language explanation: "The assertion: The product protein from gene ENSG00000103197 interacts through a particular pathway with a product of gene ENSG00000117020."

Fonte: elaborado pelos autores com base em Ferreira e van Reisen (2023).

6 À GUIA DE CONCLUSÃO

A concretização do conceito de nanopublicação não significa o aniquilamento das publicações acadêmicas tradicionais, muito pelo contrário. As nanopublicações complementam narrativas de conhecimento científico redigidas por e para seres humanos com declarações científicas elementares, acionáveis por máquinas, simples e diretas, que estimulam o compartilhamento, descoberta de recursos, acessibilidade, citabilidade e facilitam a interoperabilidade (Pensoft.Blog, 2023), peça-chave para a imprescindível interdisciplinaridade científica. Trata-se, portanto, de uma ampliação técnica e conceitual das narrativas acadêmicas e do aprofundamento do necessário escrutínio científico que valida as suas práticas.

Diante do cenário entrópico provocado pelo aumento vertiginoso da quantidade de comunicações científicas, o modelo de publicação unificado e semântico baseado em nanopublicações pode tornar o processo de comunicação científica contemporânea mais eficaz e mais próximo da lógica

veloz de uma ciência intensiva em dados e imerso em tecnologias digitais e semânticas avançadas.

A base já instalada de servidores de nanopublicações – que as colocam na categoria de tecnologia corrente - e os estudos conduzidos em diversas dimensões, que vão desde a concepção do conceito até suas aplicações em diversos domínios, têm indicado que este novo tipo de registro e comunicação científica é um constructo mais adequado para a apresentação das relações entre dados de pesquisa e o compartilhamento eficiente de conhecimento quando comparado aos meios convencionais baseados em textos não estruturados, não integrados em rede e com uma semântica computacional difusa.

As conexões do conceito já consolidado de nanopublicação com os Princípios FAIR e constructos ainda em desenvolvimento, como *FAIR Digital Objects* (FDO) e *Internet of FAIR Data & Services*, e o alinhamento da noção de nanopublicações às tecnologias da *web* semântica e a dados ligados, criam uma singularidade na evolução da comunicação científica. Criam também, de uma forma mais abrangente, uma convergência ampla e profunda entre a Ciência da Informação e as diversas vertentes da inteligência artificial, da *blockchain* e da simulação computacional, para citar alguns tópicos relevantes. Isto nos conduz a uma agenda de pesquisa rica em novas possibilidades de pesquisas interdisciplinares que vão inspirar o prosseguimento do presente estudo.

REFERÊNCIAS

BUCUR, Cristina-Iulia; KUHN, Tobias; CEOLIN, Davide. A Unified Nanopublication Model for Effective and User-Friendly Access to the Elements of Scientific Publishing. **Lecture Notes in Computer Science**, v. 12387, p. 104-119, 2020.

FERREIRA, André Cardoso Silva; VAN REISEN, Mirjam. Hourglass-based interoperability through nanopublications in VODAN-A. **FAIR Connect**, v. 1, p. 5-11, 2023. Disponível em: <https://content.iospress.com/download/fair-connect/fc221512?id=fair-connect%2Ffc221512>. Acesso em: 10 nov. 2024.

GRAY, Alasdair J. G. *et al.* **Nanopublication Guidelines**. Nanopublications, 2024. Disponível em: https://nanopub.net/guidelines/working_draft/. Acesso em: 06 nov. 2024.

GRAY, Jim *et al.* **Scientific data management in the coming decade**. Redmont, WA: Microsoft Corporation, 2005. Disponível em <https://arxiv.org/ftp/cs/papers/0502/0502008.pdf>. Acesso em: 19 jul. 2024.

GROTH, Paul; GIBSON, Andrew; VELTEROP, Jan. The anatomy of a nanopublication. **Information Services & Use**, v. 30, 2010. Disponível em: <https://content.iospress.com/download/information-services-and-use/isu613?id=information-services-and-use%2Fisu613>. Acesso em: 22 out. 2024.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. Jim Gray on eScience: A transformed scientific method. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (ed.). **The fourth paradigm: Data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. Disponível em <http://itre.cis.upenn.edu/myl/JimGrayOnE-Science.pdf>. Acesso em: 29 jul. 2024.

KUHN, Tobias; BARBANO, Paolo Emilio, NAGY, Mate Levent; KRAUTHAMMER, Michael. Broadening the scope of nanopublications. **Lecture Notes in Computer Science**, v. 7882, p. 487–501, 2013.

KUHN, Tobias; DUMONTIER, Michel. Genuine semantic publishing. **Data Science**, v. 1, 2017. Disponível em: <http://itre.cis.upenn.edu/myl/JimGrayOnE-Science.pdf><https://content.iospress.com/download/data-science/ds010?id=data-science/ds010> . Acesso em: 22 out. 2024.

MONS, Barend *et al.* The value of data. **Nature Genetics**, v. 43, n. 4, 2011. Disponível em: <https://www.nature.com/articles/ng0411-281>. Acesso em: 22 out. 2024.

MONS, Barend; VELTEROP, Jan. Nano-publication in the e-science era. *In*: WORKSHOP ON SEMANTIC WEB APPLICATIONS IN SCIENTIFIC DISCOURSE (SWASD 2009). 2009. Disponível em: [https://www.w3.org/wiki/images/4/4a/HCLS\\$\\$ISWC2009\\$\\$Workshop\\$Mons.pdf](https://www.w3.org/wiki/images/4/4a/HCLS$$ISWC2009$$Workshop$Mons.pdf) Acesso em: 22 out. 2024.

NANOPUBLICATIONS. **What is a Nanopublication?** 2022. Disponível em: <https://nanopub.net/>. Acesso em: 11 nov. 2024.

PAULA, Lorena Tavares de; MOURA, Maria Aparecida. Nanopublicações e indexação: processos e relações. **Informação & Informação**, Londrina, v. 20, n. 1, p. 137-161, 2015.

PENSOFT.BLOG. **Nanopublications tailored to biodiversity data**. 2023. Disponível em: <https://blog.pensoft.net/2023/09/12/nanopublications-tailored-to-biodiversity-data/>. Acesso em: 11 nov. 2024.

PÉREZ-GONZÁLEZ, Lourdes. Modelo/s de coste para la preservación de los datos científicos em la e-ciencia. *In*: JORNADAS DE GESTIÓN DE LA INFORMACIÓN, 12., 2010, Madrid. **Anales [...]**. Madrid: SEDIC, 2010.

Disponível em: <http://eprints.rclis.org/8555/1/Perez.pdf>. Acesso em: 06 jul. 2024.

SALES, Luana Farias; SAYÃO, Luís Fernando. Conectando a eScience à Ciência da Informação: o big metadado científico e suas funcionalidades. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 21, n. 00, p. e023017, 2023. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8673740>. Acesso em: 22 out. 2024.

SANTOS, Luiz Olavo Bonino da Silva *et al.* Towards a conceptual model for the FAIR Digital Object Framework. 23 Feb. 2023. Disponível em: <https://arxiv.org/pdf/2302.11894.pdf>. Acesso em: 20 jun. 2023.

SAYÃO, Luís Fernando; SALES, Luana Farias; CARVALHO SEGUNDO, Washington Luís R. de. Objeto Digital Fair: ampliando as fronteiras da gestão de dados. **Informação & Sociedade**, João Pessoa, v. 33, 2024. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/68646>. Acesso em: 23 out. 2024.

SCHULTES, Erik Anthony. FAIR digital object for academic publishers. **Information Services & Uses**, v. 44, 2024. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu230227>. Acesso em: 22 out. 2024.

SCHULTES, Erik Anthony *et al.* The Comparative Anatomy of Nanopublications and FAIR Digital Objects, 2022. Disponível em: <https://riojournal.com/article/94150/element/4/8025864//>. Acesso em: 22 out. 2024.

SILVA, Romário Laltany G. da; BORDIN, Andréa Sabedra. Método para criação e recuperação de nanopublicações uma aplicação no domínio de Análise de Redes de Colaboração Científica. *In*: BRAZILIAN E-SCIENCE WORKSHOP (BRESCI), 14., 2020, Cuiabá. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 73-80. Disponível em: <https://sol.sbc.org.br/index.php/bresci/article/view/11184/11055>. Acesso em: 22 out. 2024.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 22 mar. 2023.

WOUTERS, Paul. What is the matter with e-science? – thinking aloud about informatisation in knowledge creation. **Pantaneto Forum**, n. 23, July 2006. Disponível em: <http://pantaneto.co.uk/OLD%20site/issue23/wouters.htm>. Acesso em: 06 jan. 2023.

NANOPUBLICATIONS: KNOWLEDGE REPRESENTATION APPROACHES AS AN INNOVATION FOR SCIENTIFIC COMMUNICATION

ABSTRACT

Objective: this study aims to present the notion of nanopublication, a construct presented by the Concept Web Alliance, whose concept is based on the division of knowledge into statements with provenance and context, becoming a minimum publishable and citable unit, structured in RDF and other semantic technologies that allow interpretation by humans and actionability by machines. **Methodology:** the study is exploratory and bibliographic in nature and is based on the authors and institutions that support the proposal. **Conclusion:** The analysis carried out allows us to conclude that the creation of nanopublication servers and networks in conjunction with more recent concepts, such as FDO and IFDS, has the potential to change the scenario of scientific communication, constituting less a rupture of the classic model and more a technical and conceptual expansion necessary for the dynamics of eScience. **Conclusion:** It is concluded that the adoption of nanopublications and concepts such as FDO and IFDS offers an essential evolution for scientific communication in eScience, promoting a more efficient and interoperable structure between humans and machines

Descriptors: Nanopublication. Semantic technologies. Scientific communication.

NANOPUBLICACIONES: ENFOQUES DE REPRESENTACIÓN DEL CONOCIMIENTO COMO INNOVACIÓN PARA LA COMUNICACIÓN CIENTÍFICA

RESUMEN

Objetivo: este estudio tiene como objetivo presentar la noción de nanopublicación, un constructo presentado por la Concept Web Alliance cuyo concepto se basa en la división del conocimiento en declaraciones con procedencia y contexto, convirtiéndose en una unidad mínima publicable y citable, estructurada en RDF y otras tecnologías semánticas que permiten la interpretación por humanos y la accionabilidad por máquinas. **Metodología:** el estudio es de carácter exploratorio y bibliográfico y se basa en los autores e instituciones que apoyan la propuesta. **Resultado/conclusión:** El análisis realizado permite concluir que la creación de servidores y redes de nanopublicaciones, en conjunto con conceptos más recientes como FDO e IFDS, tiene el potencial de cambiar el escenario de la comunicación científica, constituyéndose menos en una ruptura del modelo clásico y más en una expansión técnica y conceptual necesaria para las dinámicas de la eScience. **Conclusion:** It is concluded that the adoption of nanopublications and concepts such as FDO and IFDS offers an essential evolution for scientific communication in eScience, promoting a more efficient and interoperable structure between humans and machines

Descriptores: Nanopublicación. Tecnologías semánticas. Comunicación científica.

AGRADECIMENTOS

Ao CNPq e à FAPERJ pelo financiamento dessa pesquisa, através de nossas bolsas de produtividade 312627/2023-8 e 313188/2023-8 e edital Jovem Cientista do estado do Rio de Janeiro E26/201.380/2022.

À Teodora Marly Gama pela cuidadosa revisão.

Recebido em: 15.11.2024

Aceito em: 22.11.2024