

SUELY LUCENA CANHA

**LINGUAGENS DE MARCAÇÃO E PERIÓDICOS CIENTÍFICOS: evolução ao
padrão Journal Article Tag Suite - JATS**

São Paulo

2021

SUELY LUCENA CANHA

**LINGUAGENS DE MARCAÇÃO E PERIÓDICOS CIENTÍFICOS: evolução ao
padrão Journal Article Tag Suite - JATS**

Trabalho de Conclusão de Curso
apresentado, como exigência parcial para a
obtenção do título de Bacharela em
Biblioteconomia, ao Departamento de
Informação e Cultura da Escola de
Comunicações e Artes da Universidade de
São Paulo

Orientador: Prof. Dr. Rogério Mugnaini

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Canha, Suely Lucena

LINGUAGENS DE MARCAÇÃO E PERIÓDICOS CIENTÍFICOS: evolução ao padrão Journal Article Tag Suite - JATS / Suely Lucena Canha; orientador, Rogério Mugnaini. -- São Paulo, 2021. 69 p.: il.

Trabalho de Conclusão de Curso (Graduação) - Departamento de Comunicações e Artes/Escola de Comunicações e Artes /Universidade de São Paulo.

Bibliografia

Versão corrigida

1. Periódicos científicos 2. Linguagens de marcação
I.Mugnaini, Rogério II. Título.

CDD 21.ed. - 020

Folha de aprovação

Nome: Suely Lucena Canha

Título: LINGUAGENS DE MARCAÇÃO E PERIÓDICOS CIENTÍFICOS: evolução ao padrão Journal Article Tag Suite - JATS

Aprovado em: 25 /02/ 2021

Banca:

Nome: Francisco Carlos Paletta

Instituição: Universidade de São Paulo, Escola de Comunicações e Artes.

Nome: Amanda de Souza Ramalho

Instituição: Scientific Electronic Library Online - SciELO

Agradecimentos

A Ele, que em suas manifestações mais sutis mostra que a vida pode ser leve e bela.

A todos os professores com quem tive aula ao longo desses quatro anos de graduação em Biblioteconomia na ECA-USP, pelo empenho e dedicação com que compartilharam seus conhecimentos. Em especial a meu professor orientador Rogério Mugnaini, cujas valiosas contribuições foram essenciais para a elaboração deste trabalho.

A minha família, João Paulo e Simone e especialmente à Rosa Lucena Canha, minha irmã, pela preocupação, atenção, companheirismo e por sempre estar presente em todos os momentos me apoiando e incentivando.

Aos membros do SciELO Brasil: Solange Santos, Ednilson Gesseff e Marina Chagas, pela disponibilidade e conversas descontraídas que foram tão importantes para a redação e desenvolvimento deste trabalho. Agradeço também à Francine Currivil, Carolina Taniguchi, Luana Araújo e Isabela Gagliardi, pela compreensão e mensagens de apoio.

***Aprender é a única coisa de que a mente nunca
se cansa, nunca tem medo e nunca se arrepende***

(Leonardo da Vinci)

RESUMO

Este trabalho aborda a utilização das linguagens de marcação SGML, HTML e XML com destaque para o XML-JATS, nos processos de produção e publicação de periódicos científicos eletrônicos. Sob uma perspectiva descritiva e exploratória, a elaboração do estudo buscou investigar primeiramente o surgimento dos primeiros periódicos e a transição da publicação impressa para a eletrônica. Em seguida, foi realizada breve caracterização sobre aspectos gerais do SGML, HTML e XML e as razões para seu aprimoramento orientado às publicações científicas. A metodologia adotada consistiu em revisão da literatura sobre os dois modelos de publicação supracitados, o surgimento de linguagens pensadas para as publicações científicas e estudo de documentações técnicas e casos de uso do padrão XML-JATS. As considerações finais, - embora este tema tenha sido pouco explorado em estudos nacionais a partir da perspectiva da biblioteconomia - demonstram que a utilização de linguagens de marcação desenvolvidas especificamente para periódicos eletrônicos sugere melhorias significativas nos processos de descrição, armazenamento, intercâmbio e manuseio de informações entre bases eletrônicas de conteúdos científicos.

Palavras-chave: Periódicos científicos. Linguagens de marcação. JATS-XML. Periódicos eletrônicos.

ABSTRACT

This work addresses the use of the markup languages SGML, HTML and XML emphasizing XML-JATS, in the production and publication processes of electronic scientific journals. Under a descriptive and exploratory perspective, the study preparation firstly sought to investigate the emergence of the first journals and their transition from print to electronic publication. Following, there is a brief characterization on the general aspects of SGML, HTML and XML and the reasons for its improvement oriented to scientific publications. The adopted methodology consists of a literature review on the two aforementioned publication models, the emergence of languages designed for scientific publications and the study of technical documentation and use cases of XML-JATS standard. The final considerations – although this theme has been little explored in national studies from the library science perspective – demonstrate that the use of markup languages, developed specifically for electronic journals, suggest significant improvements in the processes of description, storage, exchange and handling of information between electronic bases of scientific content.

Keywords: Scientific journals. Markup languages. JATS-XML. Electronic journals.

LISTA DE FIGURAS

Figura 1 - Os dois primeiros periódicos científicos no século XVII. Journal Des Scavans e o Philosophical Transactions	17
Figura 2 – Exemplo de informação marcada em HTML	26
Figura 3 - Informação estruturada em XML com o uso de atributos	30
Figura 4 - Exemplo de especificações de um arquivo .DTD	31
Figura 5 - Exemplo de DTD declarada na tag DOCTYPE de um arquivo XML	32
Figura 6 – Página inicial da documentação da Journal Article Tag Suite	42
Figura 7 - Tag library do conjunto Journal Publishing.....	43
Figura 8 - Diagrama da estrutura básica da Archiving and Interchange Tag Set	47
Figura 9 - Diagrama da estrutura básica da Archiving Interchange Tag Set	49
Figura 10 - Diagrama da estrutura básica da Article Authoring Tag Set.....	50
Figura 11 - Estrutura geral de artigo científico.....	51
Figura 12 - Dados sobre o periódico estruturados em XML-JATS	52
Figura 13 - Exemplo de dados de autoria estruturado em JATS	53
Figura 14 - Parágrafo de artigo marcado em XML	53
Figura 15 - Exemplo de tabela marcada em XML-JATS	54
Figura 16 - Referência bibliográfica estruturada em XML-JATS.....	55
Figura 17 - Marcação de dados de financiamento.....	56

LISTA DE TABELAS

Tabela 1 - Compatibilidade entre elementos da três tag sets	44
--	----

SUMÁRIO

1. INTRODUÇÃO	9
3. OBJETIVOS	12
3.1 Objetivo geral	12
3.2 Objetivos específicos.....	12
4. METODOLOGIA.....	13
5. A COMUNICAÇÃO CIENTÍFICA.....	15
5.1 Periódicos científicos: surgimento e evolução	16
5.2 Periódicos científicos eletrônicos.....	18
6. LINGUAGENS DE MARCAÇÃO	24
6.1 Standard Generalized Markup Language - SGML	24
6.2 HyperText Markup Language – HTML.....	25
6.3 eXtensible Markup Language – XML.....	27
6.4 Document Type Definition - DTD.....	30
6.5 As linguagens de marcação e os periódicos científicos	33
6.5.1 AAP Electronic Manuscript Standard - AAP DTD	33
6.5.2 ISO 12083 - Serial DTD.....	34
6.5.3 Modular Application for journals - MAJOUR.....	35
6.5.4 Simplified SGML for Serial Headers - SSSH DTD	35
6.5.5 PMC Document Type Definition – PMC DTD	36
6.5.6 Journal Archive and Interchange Tag Suite - NLM DTD.....	38
7. ANSI/NISO Z39.96-2012 - JOURNAL ARTICLE TAG SUITE – JATS	40
7.1 Características das Tag sets - JATS	41
7.1.1 Journal Archive e Interchange Tag Set	46
7.1.2 Journal Publishing Tag Set.....	48
7.1.3 Article Authoring Tag Set.....	49
7.2 Análise de um artigo em formato XML-JATS	51
7.3 Como o padrão XML-JATS está sendo utilizado.....	57
8. CONCLUSÃO.....	61
9. REFERÊNCIAS BIBLIOGRÁFICAS.....	64

1. INTRODUÇÃO

Desde seu surgimento no século XVII, os periódicos científicos representam um importante canal de comunicação do conhecimento produzido pela comunidade científica. Sua importância se dá pelo registro, difusão e compartilhamento de descobertas e resultados de pesquisas, que contribuem para o fortalecimento da ciência e geração de novos conhecimentos.

A partir da década de 1960, a evolução da Internet tem se mostrado um meio amplamente disponível para acessar e compartilhar informações eletronicamente. De modo a acompanhar essa tendência, a publicação eletrônica de periódicos se expandiu, assim como a necessidade de novas ferramentas tecnológicas e metodologias que favorecessem tanto a produção como a difusão dos conteúdos.

A ruptura com o modelo impresso em prol das formulações digitais abriu possibilidades extraordinárias para o mundo da comunicação científica, libertando definitivamente as publicações acadêmicas dos limites bidimensionais e autocontidos do texto, inaugurando novas formulações de apresentação e interoperabilidade, e, sobretudo, estabelecendo novos padrões de cooperação e interatividade em favor da geração de novos conhecimentos (SAYÃO, 2010, p. 69).

Nesse sentido, a inserção de novas tecnologias da informação e comunicação modificou os fluxos informacionais, promovendo inúmeras possibilidades quanto ao tratamento e manejo da informação científica, fazendo com que sua abordagem no contexto atual da Biblioteconomia, enquanto área dedicada ao tratamento da informação, seja de suma importância.

O percurso evolutivo da transição do modelo impresso para o eletrônico passou por diversas fases, que vão desde o escaneamento de versões impressas até sua produção completa em ambiente digital, na qual a utilização de linguagens de marcação teve fundamental papel nos processos de produção e publicação dos conteúdos científicos.

As linguagens de marcação passaram a ser utilizadas por editores e agregadores de periódicos desde 1980 e desde então vem sendo modificadas e aprimoradas a este propósito, tornando-se cada vez mais presentes em seus fluxos editoriais.

Este trabalho aborda a utilização do SGML, HTML e o XML nos processos de produção e publicação de periódicos. Embora os dois primeiros formatos tenham sido

amplamente utilizados para esta finalidade, o XML possui destaque na literatura enquanto recurso tecnológico utilizado para reduzir algumas limitações técnicas de seus dois antecessores, sendo este o formato base para as emergentes linguagens desenvolvidas para periódicos.

A respeito das novas possibilidades a partir do uso de computadores e linguagens de marcação na produção de conteúdos científicos Van Houweling (1994, *apud* KANE E MAKUCH, 1998. p. 9, tradução nossa) ressalta que:

Novas ferramentas são necessárias para novos tempos. Já que o mundo do conhecimento muda tão continuamente e tão rapidamente, sempre há novos desafios, novas informações para serem buscadas, novos processos a serem compreendidos.

Discussões no âmbito da comunicação científica incluem questões sobre uso de metodologias mais adaptáveis e aprimoradas à tendência digital, evidenciando que editores, bibliotecas digitais e profissionais da informação envolvidos, vem absorvendo recursos tecnológicos em prol do tratamento da informação científica.

Nos últimos anos, iniciativas de profissionais da tecnologia e informação vem expandindo seus esforços a partir das linguagens de marcação, em especial o Extensible Markup Language - XML, para incluir novas potencialidades à publicação eletrônica de periódicos. Esses esforços, visam suprir demandas específicas inerentes a comunicação em rede, tais como a necessidade de dados estruturados para a recuperação, arquivamento e interoperabilidade de informações entre bases distintas, assim como maior rapidez nas publicações.

Indo ao encontro dessas demandas, pontua-se neste trabalho o surgimento em 2012 do ANSI/NISO Z39.96.2012, também conhecido como *Journal Article Tag Suite – JATS*, que consiste em três modelos de XML projetados para descrição e publicação, arquivamento e intercâmbio e controle de autoria de artigos científicos.

Mesmo as linguagens de marcação baseadas em SGML e XML principalmente, já estejam presentes na publicação científica de forma bastante consolidada, muitos editores estão aderindo a este novo padrão para otimização de processos internos de manejo de dados para finalidade e projetos específicos. Do mesmo modo, grandes bases e bibliotecas eletrônicas o tem tomado como um requisito técnico de indexação.

2. JUSTIFICATIVA

Os periódicos científicos são reconhecidamente importantes fontes de informação que contribuem para o avanço da ciência e produção de novos conhecimentos ao longo do tempo. Assim como ocorreu na sociedade em geral, eles não se mantiveram alheios aos avanços e transformações tecnológicas que impactam nas formas de produzir e disseminar informações (MIRANDA; PEREIRA, 1996; OLIVEIRA, 2008; STUMPF 1996).

Desde que se consolidaram no formato eletrônico, buscam por procedimentos e metodologias que otimizem os processos editoriais. A inserção das linguagens de marcação foi um grande avanço nesses processos, promovendo novas possibilidades informacionais e modificando a forma como são produzidas e disseminadas as informações científicas na web (BARNES, 1997; KANE e MAKUCH, 1988).

Considerando a importância da adaptabilidade da Biblioteconomia às atuais tendências editoriais de periódicos eletrônicos segundo pontado por Entralgo-Amaro; Aguiar-Cedeño (2009) e Targino (1999), este trabalho justifica-se em buscar um olhar mais especializado da área em relação às transformações tecnológicas que impactam diretamente sobre os modos de produzir e publicar conteúdos científicos.

Aborda-se a utilização das linguagens de marcação, em especial o formato XML-JATS, por se tratar de um padrão relativamente novo, baseado em XML e projetado especificamente para periódicos eletrônicos, que vêm sendo requerido por diversas bases e bibliotecas eletrônicas como requisito técnico para indexação, assim como vem sendo utilizado para otimização de diversos processos de tratamento informacional no âmbito das instituições que fazem uso dele.

3. OBJETIVOS

3.1 Objetivo geral

As linguagens de marcação tiveram importante papel no modo de produzir e publicar periódicos científicos nas últimas décadas. Assim, o objetivo geral deste trabalho é investigar algumas das principais linguagens de marcação que foram orientadas a este propósito e as razões pelas quais editores e agregadores passaram a utilizá-las nos processos editoriais de produção e publicação de periódicos.

3.2 Objetivos específicos

Num contexto de rápidas transformações tecnológicas e busca por novos procedimentos e metodologias de tratamento informacional para períodos eletrônicos a partir das linguagens de marcação, emergiu em 2012 um novo padrão, denominado ANSI/NISO Z39.96-2012, também conhecido por JATS. Ele foi criado e vem sendo aprimorado para a marcação em formato XML de periódicos científicos, por esta razão os objetivos principais deste trabalho são:

- Compreender o que é ANSI/NISO Z39.96-2012, quais suas funções e principais características;
- Investigar de que forma este padrão vem sendo utilizado em processos de publicação de periódicos eletrônicos.

4. METODOLOGIA

Este trabalho apresenta-se como uma pesquisa descritiva e exploratória que aborda o processo evolutivo dos periódicos científicos, desde seu surgimento até o presente formato eletrônico. São pontuados, além da migração de suporte, o uso de novos recursos computacionais com destaque para a inserção das linguagens de marcação, que tiveram papel fundamental nos modos de produzir e publicar literatura científica nas últimas décadas.

O levantamento bibliográfico deste trabalho foi realizado em bases nacionais e estrangeiras, foram selecionados artigos nos idiomas português, inglês e espanhol na BRAPCI, SCIELO e WOS. Fontes não localizadas nessas bases, mas que foram essenciais para a elaboração da pesquisa, foram coletadas via Google Scholar, sendo estas: dissertações, identificadas em repositórios institucionais da UFscar, UNESP e UNICAMP; anais de congresso sobre o uso do padrão XML-JATS e documentações técnicas disponibilizadas nos sites SciELO e NCBI.

Para a elaboração dos capítulos sobre a evolução dos periódicos foi realizada abordagem descritiva, baseada em revisão de literatura de artigos e trabalhos acadêmicos acerca deste tema, através dos termos de busca, “comunicação científica” e “periódicos eletrônicos”. A seleção das fontes considerou produções que tratassem primeiramente de aspectos históricos sobre o surgimento dos periódicos científicos, a transição do suporte impresso ao digital até a consolidação das atuais tendências editoriais que priorizam as publicações em formato eletrônico.

O capítulo sobre as linguagens de marcação possui dois enfoques distintos. O primeiro sob uma abordagem descritiva, trata sobre os aspectos gerais e as principais características do SGML, HTML e XML. O segundo, busca investigar quais foram as principais razões para o desenvolvimento e como se deu a utilização dessas linguagens nos processos editoriais de produção e publicação de periódicos.

Para o primeiro enfoque, foram utilizados artigos científicos e trabalhos acadêmicos pesquisados nas bases brasileiras. Dada a escassez de trabalhos nacionais que abordam este tema a partir do segundo enfoque, o levantamento bibliográfico foi realizado na Base Periódicos Capes, através de buscas por termos como XML, HTML, SGML e Eletronic Publisher, na qual recuperou-se literatura pertinente.

Para elaboração do capítulo dedicado ao padrão ANSI/NISO Z39.96-2012, o formato XML-JATS e casos de uso, realizou-se análise da sua documentação oficial, manuais técnicos denominados *Tag Libraries* e trabalhos apresentados em conferências que descrevem alguns casos concretos sobre a utilização do XML-JATS.

A construção dos exemplos de estruturas de XML considerou a análise das *Tag Libraries* da JATS, e foi utilizado o *Oxygen Xml Editor*. Um software desenvolvido para a criação, edição e validação de arquivos em formato .XML.

5. A COMUNICAÇÃO CIENTÍFICA

A comunicação científica engloba atividades associadas à produção, disseminação e uso da informação, desde o momento em que uma ideia se torna objeto de pesquisa de um cientista, até a aceitação de seus resultados como constituinte do estoque universal de conhecimento (MIRANDA; PEREIRA, 1996).

O processo de comunicação científica caracteriza-se pela investigação, análise, documentação, comunicação, produção, registro e disseminação da informação. A fragmentação do conhecimento em áreas cada vez mais especializadas das ciências contribuiu para delimitar e legitimar novas disciplinas e campos de estudos, que se materializaram na forma de periódicos.

Isto significou a consolidação dos periódicos enquanto formalização do processo de comunicação, permitindo que as pesquisas, uma vez registradas pudessem manter-se acessíveis e compartilhadas entre pares, sendo de fundamental importância para produção de novos conhecimentos.

Miranda e Pereira (1996) destacam duas principais funções dos periódicos: a de comunicação do conhecimento à sociedade e a comunicação entre os pares de um determinado campo do saber. Nesses processos estão aspectos tanto sobre a visibilidade, quanto a credibilidade das produções, que servem ao fortalecimento e desenvolvimento da ciência.

Kane e Makuch (1998), atribuem à “explosão informacional” que caracterizou a década de 1960, o marco fundamental para o aumento de periódicos em formato eletrônico, do mesmo modo impulsionou a busca por novas maneiras de organização e armazenamento, que viabilizassem o acesso e a disseminação das informações científicas de forma mais ampla e rápida.

Desde então, inovações tecnológicas têm transformado a forma como os conteúdos científicos são produzidos e a maneira que os profissionais da informação passaram a lidar com esse volume informacional. Fez-se crescente a necessidade tanto de infraestruturas de gerenciamento, assim como novas ferramentas e procedimentos que assegurem o acesso, compartilhamento e visibilidade das pesquisas ao longo do tempo. Isso significa que os periódicos em suas funções primordiais, foram diretamente impactados pelas transformações e avanços

tecnológicos computacionais, característicos da sociedade contemporânea. Mediante este contexto, práticas e metodologias mais adaptáveis ao ambiente digital passaram a estar cada vez mais presentes nos fluxos editoriais (MIRANDA; PEREIRA, 1996).

5.1 Periódicos científicos: surgimento e evolução

Os periódicos científicos representam um importante canal formal de comunicação do conhecimento gerado pela comunidade científica e desde seu surgimento no século XVII, expressam papel fundamental na difusão e compartilhamento de descobertas e resultados de pesquisa.

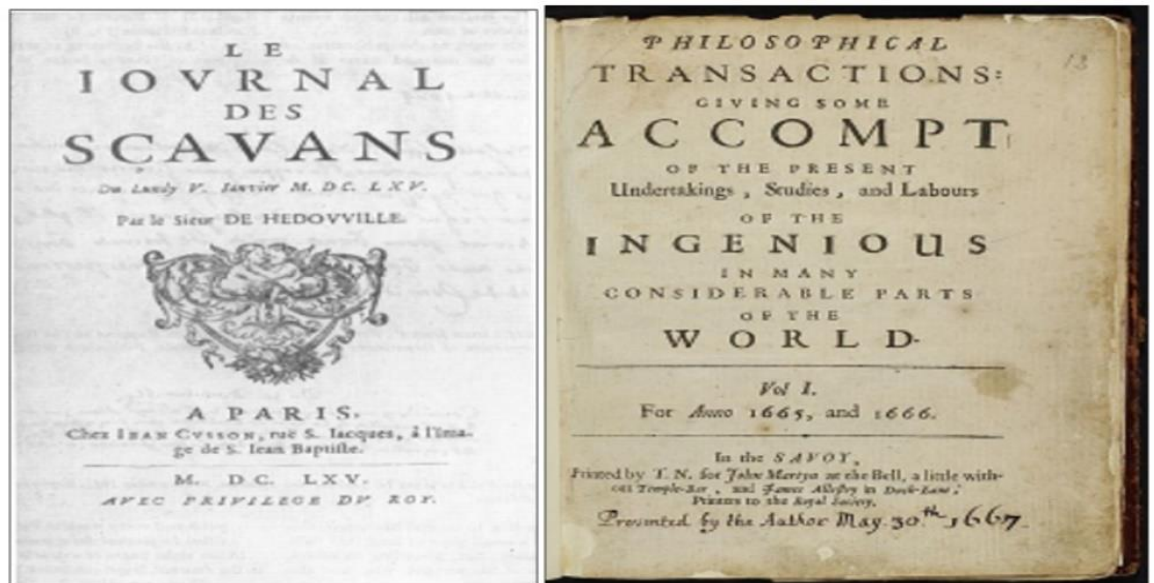
Segundo Stumpf (1996), os dois primeiros periódicos científicos datam de 1665. Foram eles: o *Journal des Sçavants* e o *Philosophical Transactions da Royal Society of London* (Figura 1). Para a autora, eles contribuíram na consolidação de modelos distintos de publicação de literatura científica.

O periódico francês influenciou o desenvolvimento das revistas dedicadas à ciência geral, sem comprometimento com uma área específica. Sua produção consistia em geral na catalogação e resumo de importantes livros publicados à época na Europa, relatórios técnicos e científicos sobre descobertas nos campos da ciência e das artes. Em 1816, o *Journal des Sçavants* consolidou-se como periódico literário e passou a ser publicado pela Academia de Letras Francesas.

O *Philosophical Transactions da Royal Society of London* contemplava a divulgação de experimentos científicos e cartas de membros das sociedades e academias científicas na Europa durante o século XVIII. Ele é considerado como precursor do periódico científico moderno, dado ao fato de ter sido pioneiro nos processos de revisão por pares dos artigos produzidos pelas sociedades científicas.

No Brasil, a fundação da Imprensa Régia, no ano de 1808, propiciou o surgimento de duas publicações que são consideradas como os primeiros periódicos científicos brasileiros: a Gazeta Médica do Rio de Janeiro, no ano de 1862 e a Gazeta Médica da Bahia, em 1866 (SOUZA *et al.*, 2002).

Figura 1 - Os dois primeiros periódicos científicos no século XVII. Journal Des Scavans e o Philosophical Transactions



Fonte: <https://blog.scielo.org/blog/2015/03/05/350-anos-de-publicacao-cientifica-desde-o-journal-des-scavans-e-philosophical-transactions-ate-o-scielo/#.X4-RHohKhHY>

A partir do século XIX, a produção de periódicos científicos cresceu significativamente face aos avanços das tecnologias de impressão e fabricação de papel, que asseguravam o registro das produções científicas e acadêmicas. Por ser um tipo de publicação periódica, tem como característica a rápida divulgação de resultados e novos achados científicos.

A partir da segunda metade do século XX, o número de periódicos científicos passou a ter um crescimento exponencial, como nunca visto, dado ao fato de passarem a ser publicados por editores comerciais e universidades, e principalmente pela introdução do uso de computadores nos processos editoriais de produção e veiculação (STUMPF, 1996).

5.2 Periódicos científicos eletrônicos

Desde o surgimento dos primeiros periódicos no século XVII, significativas transformações puderam ser observadas em sua história através das tecnologias de comunicação e informação TICs, tanto em relação aos modos de produzi-los e disseminá-los quanto ao manejo e uso das informações registradas.

Uma das transformações mais significativas nas publicações de periódicos foi a transição do seu formato impresso para publicações criadas em ambiente eletrônico, permitindo que o acesso à informação ocorresse de maneira muito mais rápida.

Nas últimas décadas do século XX, e mais intensamente nos primeiros anos do século sucessor, as facilidades proporcionadas com o uso das tecnologias digitais fizeram com que os periódicos impressos migrassem para o formato eletrônico de forma muito rápida. Mesmo com essa migração, alguns permaneceram durante bom tempo com o formato impresso, caracterizando as versões híbridas. No entanto, nos últimos anos, observa-se um crescimento e uma tendência cada dia mais frequente, do abandono da versão impressa, condicionando a disponibilização apenas do formato eletrônico (SANTA ANNA, 2018, p. 71).

Entende-se, portanto, que o crescente número de publicações em formato eletrônico favoreceu os propósitos de difusão e divulgação de pesquisas produzidas pelas comunidades científicas em todo o mundo.

Segundo Stumpf (1996), a terminologia “*Publicação Eletrônica*” surgiu quando os periódicos científicos passaram a utilizar e explorar recursos computacionais e eletrônicos como hipertexto e hipermídia, tanto nas etapas de criação quanto de armazenamento e disseminação, porém esse processo ocorreu paulatinamente.

No final da década de 1970, quando a internet e as tecnologias e ferramentas computacionais permitiram expressivo avanço nos processos de editoração eletrônica, a terminologia ganha contornos mais precisos e passa a abarcar dois modos distintos de publicação: os “periódicos on-line” e os “periódicos eletrônicos”.

Para Lancaster (1995), essa distinção se faz necessária, pois reflete diretamente nas características intrínsecas do suporte informacional do periódico, do momento de sua concepção até a sua disponibilização. O autor sintetiza a evolução dos periódicos eletrônicos em quatro etapas principais, cada uma delas influenciando em níveis diferentes a forma como os conteúdos informacionais passavam a ser tratados pelos serviços de informação.

A primeira fase, se refere à época em que computadores passaram a ser utilizados na produção dos periódicos, ainda que as publicações se mantivessem no formato impresso. Na segunda etapa, as publicações eram disseminadas em forma eletrônica. Isto é, elas passaram ao meio digital via recursos como o CD-ROM, microformas ou escaneamento, mas mantinham-se cópias fiéis de sua versão impressa. A partir da terceira fase os periódicos passaram a exibir um formato diferenciado e alguns recursos a mais foram agregados. Sendo descritas, como o momento em que o periódico passa a explorar as possibilidades eletrônicas de maneira mais ampla (LANCASTER, 1995).

Desta forma, compreende-se que o conceito de “*periódico on-line*” remete aos periódicos impressos que passaram ao meio digital via recursos computacionais usados para a migração de suporte. Já o conceito de “*periódicos eletrônicos*” remete a ideia de periódico criado para o meio eletrônico e disponível apenas nesse formato, que já eram criados contando com recursos até então inaplicáveis a versão impressa.

Dentre os projetos pioneiros que se dedicaram a criação dos primeiros periódicos eletrônicos, destacam-se o norte americano “*Electronic Information Exchange System*”, desenvolvido pelo New Jersey Institute Of Technology em 1978-1980 e o Birmingham and Loughborough Electronic Network Development (BLEND) que desenvolveu o periódico *Computer Human Factors*. Estas iniciativas engajaram-se no desenvolvimento de processos de melhorias que pudessem acelerar a produção automatizada de periódicos apenas em formato eletrônico (TARGINO, 1999).

No Brasil, o primeiro periódico eletrônico foi o *Journal of Venomous Animals and Toxins*. Lançado em 1995, foi desenvolvido pela Universidade Estadual Paulista (Unesp) e mantém suas publicações até a atualidade. Suas primeiras edições foram realizadas em disquete, porém, atualmente existe apenas em formato eletrônico (BERTIN, 2008, *apud* CORREA, 2019).

A evolução tecnológica permitiu um progresso substancial nas questões relativas ao suporte informacional e uso da informação, pois computadores têm se tornado cada vez mais poderosos, mais rápidos, menos caros e mais fáceis de usar. Tornando-se ferramentas cada vez mais indispensáveis para lidar com a enorme quantidade de informações contidas em vários tipos de documentos, inclusive os científicos (KANE; MAKUCH, 1998).

[...] pode-se considerar que a utilização da tecnologia de informação e comunicação (TIC) já está incorporada a praticamente toda cadeia de comunicação, pois além dos estágios iniciais (informais), todos os outros passos já se realizam em ambiente virtual (OLIVEIRA, 2008, p.69).

A publicação de periódicos eletrônicos cresceu exponencialmente principalmente face ao surgimento e aos avanços da internet que tornaram o ritmo geral da sociedade mais rápido e o volume informacional cada vez mais significativo. Para Kane e Makuch (1998), o fator que levou as organizações envolvidas na comunicação científica a publicarem eletronicamente foi a quantidade cada vez maior de documentos para publicar, gerenciar e tornar disponível.

No atual momento dos periódicos, ainda que muitos se mantenham híbridos, sua concepção de criação e desenvolvimento são pensadas para o ambiente digital. Além da maior rapidez na produção e disponibilização, o formato eletrônico conta com incrementos impossíveis de serem aplicados em suportes impressos e abrem inúmeras possibilidades de acesso e uso da informação. Isso significa um importante elemento de dinamização na comunicação entre pesquisadores e a sociedade, assim como contribui para melhorias na acessibilidade e visibilidade das pesquisas científicas em larga escala.

Contudo, algumas problemáticas permearam o início do percurso de transição do modelo de publicação, constituindo percalços que precisaram ser superados até que fosse possível alcançar as potencialidades, que atualmente podem lograr os periódicos eletrônicos.

A princípio, parte da comunidade científica, que envolvia autores, editores e universidades, mostrava certa resistência ao modelo eletrônico de publicação. Essencialmente, essa oposição remetia à aspectos sobre a credibilidade do conteúdo e a permanência dos recursos digitais ao longo do tempo.

Qualquer mudança requer aceitação por parte da maioria da comunidade científica envolvida. Atualmente, a posição das publicações eletrônicas ainda não se encontra totalmente consolidada. Algumas universidades ainda não aceitam que as publicações eletrônicas sejam equivalentes às impressas em papel no que concerne a certas finalidades como, por exemplo, na decisão quanto à promoção de um docente. Alguns editores de periódicos vêem a publicação na Internet como semelhante a uma notícia veiculada pela mídia. Não aceitarão subsequentemente o mesmo material para publicação em suas revistas impressas em papel e tampouco o considerarão como uma fonte aceitável para ser citada (MEADOWS, 1999, p. 206).

Esses fatores fizeram com que muitos periódicos mantivessem suas publicações em formato impresso, mesmo diante da iminência da migração para o formato eletrônico. Para Packer (1988, p.109) foi somente na segunda metade dos anos 90 que a publicação científica eletrônica passou a ser aceita *“universalmente como um fenômeno inexorável pela maioria dos atores do processo de comunicação científica”*.

Aspectos de caráter técnico e tecnológico também constituíram questões importantes no processo de transição dos periódicos. Dos anos 1980 ao início da década seguinte, quando se configurou a fase em que os periódicos ascendiam ao meio eletrônico por meio de tecnologias como o CD-ROM ou escaneamento, tanto os editores quanto as instituições disseminadoras de informação se depararam com situações tais como, a migração de formato e a manutenção de suportes proprietários utilizados para acessar os conteúdos que representavam altos custos, complexidade e implicações técnicas para a efetiva manutenção dos periódicos por longos períodos.

Os processos de conversão dos recursos impressos realizados por editores, ocorriam geralmente por meio de digitalização e criação de imagens bitmap¹. As bases indexadoras precisavam criar registros bibliográficos descritivos e resumos para vincular essas informações aos recursos digitalizados e assim torná-los recuperáveis nos sistemas de informação (BARNES, 1997).

Essa metodologia servia a mecanismos de recuperação da informação, mas os resultados trazidos eram normalmente generalistas e pouco precisos. Isto ocorria pois no contexto de sua aplicação, os buscadores consistiam em recuperar apenas informações intrínsecas aos documentos, isto é, ocorrência de determinados termos no texto, que nem sempre retratavam o significado esperado no contexto de busca pelos usuários.

Embora as etapas iniciais da transição para o meio eletrônico tenham representado um significativo avanço na recuperação da informação científica através

¹ O bitmap é um conjunto de pixels, também chamados de “pontos”, que define a resolução de uma imagem digital. A resolução mede em DPI, a qualidade de um bitmap. DPI é uma sigla em inglês e significa **“dots per inch”** – em português: **“pixels (ou pontos) por polegada”**. Quanto maior o DPI de uma imagem, maior sua definição e qualidade.

do uso de computadores se comparado apenas com a versão impressa, o processo ainda era longo e dispendioso (BARNES, 1997; MEADOWS, 1999).

A partir da década de 1990, com a ampliação do uso do XML e o desenvolvimento da Web semântica, os motores de busca passaram a considerar o conteúdo semântico das informações nos processos de recuperação (MORAES; SOARES, 2004).

Posteriormente, foram criados projetos para possibilitar que a inserção de periódicos em bases eletrônicas fosse realizada através do fornecimento de dados brutos de periódicos eletrônicos diretamente às bases. Esse processo consistia em remessas de arquivos de texto e imagens, sendo então as bases indexadoras responsáveis por armazenar os periódicos e desenvolver os softwares necessários para acessá-los. Dentre iniciativas dessa natureza, pontua-se o surgimento do projeto *TULIP*, iniciado em 1991 pela editora científica Elsevier, porém a consolidação de um sistema que comportasse inúmeras variações e formatos mostrou-se inviável e tão dispendiosa para os periódicos quanto os processos de migração de suporte, fazendo com que o projeto terminasse em 1995 (BARNES, 1997; GUÉDON, 2014).

Segundo Meadows (1999), a concepção de criação de periódicos para o meio digital e a inclusão de recursos de multimídia, hipertexto e links que ligavam o conteúdo dos artigos a fontes externas, representou um grande avanço. Ao enriquecerem o documento desta forma, novas experiências informacionais e de leitura eram possíveis, como das vantagens em processos editoriais de controle de qualidade e revisão por pares que passaram a contar com transversalidade da comunicação em rede.

A internet não significou apenas a ampliação ao acesso à informação pela sociedade de modo geral. Ela impactou fortemente a dinâmica informacional entre pesquisadores, favorecendo os canais de comunicação científica e a interação entre autores e pesquisadores, que passaram a trocar conhecimentos com maior rapidez e de maneira integrada (MEADOWS, 1995; TARGINO, 2000).

O desenvolvimento da Web permitiu grande expansão na comunicação, ao reduzir as barreiras de entrada para publicação eletrônica e favorecendo sua distribuição de maneira onipresente. Contudo, inúmeras outras questões relacionadas a interfaces, formatos de documentos e processamento apontavam para

necessidades de infraestruturas e conhecimentos técnicos necessários para arquivar e acessar as informações de forma mais estável, eficaz e com custos reduzidos.

Para Barnes (1997), a internet parece ter sido um problema e uma solução ao mesmo tempo, pois ao passo que as informações passaram a ser mais amplamente disponibilizadas e acessíveis, tornavam-se cada vez mais necessárias estratégias computacionais que permitissem sua descoberta em um universo informacional tão diverso. Nesse sentido, a estruturação de dados e legibilidade na web passaram a ser pontos centrais que demandavam ferramentas tecnológicas mais aprimoradas, evidenciando claramente a necessidade de reformulação dos modelos de armazenamento de dados, para posterior coleta e compartilhamento. Ainda, segundo o autor, tais possibilidades representavam o mérito inerente a web, sendo esta própria, o elemento padronizador necessário para estimular efetivamente a migração dos periódicos para o formato eletrônico.

6. LINGUAGENS DE MARCAÇÃO

As linguagens de marcação surgiram no domínio das Ciências da Computação anteriormente a sua incorporação pela Biblioteconomia nos fluxos de tratamento informacional. Elas foram pensadas como uma solução para organizar o crescente volume de informação digital que vinha sendo produzido especialmente por empresas do ramo tecnológico.

Codificar ou "marcar" um texto é um processo de indicar estrutural e semanticamente como um conteúdo informacional deve ser interpretado por algum sistema de informação automatizado. Para Almeida (2002, p. 6), a linguagem de marcação pode ser entendida como um “*conjunto de convenções utilizadas para a codificação de textos*”, na qual deve estar especificado quais marcações (tags) são permitidas, quais são exigidas, assim como sua cardinalidade e estrutura num determinado arquivo.

Segundo Bax (2000), a utilização de padrões de marcação internacionais e não proprietários, tais como o SGML, HTML e XML possibilita a criação de documentos portáteis, ou seja, documentos que não são dependentes de um determinado software, hardware ou sistema operacional para serem manipulados por sistemas.

Dentre as principais linguagens de marcação existentes, pontua-se neste trabalho algumas das características essenciais sobre o Standard Generalized Markup Language - SGML, HyperText Markup Language – HTML e o eXtensible Markup Language – XML.

6.1 Standard Generalized Markup Language - SGML

Necessidades práticas relacionadas ao manejo de documentos eletrônicos da International Business Machines (IBM) serviram para impulsionar projetos de pesquisas voltados à padronização de marcações que pudessem favorecer o processamento automatizado dos documentos eletrônicos.

Para este propósito, os pesquisadores Charles Goldfarb, Edward Mosher e Raymond Lorie se empenharam no desenvolvimento de uma linguagem computacional específica e publicaram em 1969, a primeira versão da Generalized Markup Language - GML, considerada como a precursora das linguagens de marcação computacionais.

Em 1978, representantes da linguagem GML se uniram em comitê da American National Standards Institute (ANSI) com o objetivo de desenvolver uma linguagem de marcação padrão que pudesse ser mais amplamente aplicada em documentos digitais. O objetivo principal desta ação era a representação padronizada e estruturada das informações através do uso de tags passíveis de serem processadas por sistemas computacionais, independentemente do sistema operacional ou software utilizado.

Quando normalizado sob a ISO 8879/86 em 1986, essa linguagem passa a ser denominada *Standardized Markup Language - SGML* e começa a ser utilizada em grande escala em projetos governamentais e comerciais para o compartilhamento de documentos legíveis por máquinas (Ramalho et al, 2017).

Um documento SGML em particular é configurado basicamente em quatro partes. Que são uma declaração que define opções específicas do padrão que está sendo implementada para um determinado aplicativo ou documento; uma definição de tipo de documento denominada DTD, que descreve a relação dos elementos estruturais (tags) em um tipo de documento; a instância ou texto marcado com o tags e uma folha de estilo que descreve como os elementos estruturais que devem ser apresentados em um página web (KANE; MAKUCH 1998).

Em linhas gerais, arquivos em SGML carregam consigo a especificação de sua própria gramática formal, isto significa que somente sistemas computacionais podem interpretar as marcações existentes por meio da associação de um recurso que descreve e valida estruturalmente um documento, denominado Document Type Definition (DTD).

Dada sua característica complexa e não legível por humanos, sua popularização e ampla adesão entre profissionais da informação que não eram especialistas na área de tecnologia foram limitadas. Isso fez com que diversos grupos de trabalho no âmbito da Tecnologia da Informação voltassem seus esforços para a criação de códigos abertos mais acessíveis e fáceis de manejar.

6.2 HyperText Markup Language – HTML

O HTML surgiu como uma especificação SGML feita para funcionar na internet, focando prioritariamente a apresentação das informações. Na década de 1980,

Berners-Lee criou o *HyperText Markup Language - HTML*, sendo esta uma das mais populares aplicações para SGML, que possibilitou apresentar de maneira mais simples os diferentes tipos de recursos informacionais em páginas Web.

Ele foi desenvolvido com foco na criação de páginas, considerando os seguintes aspectos: definição de elementos para construção de páginas HTML e a descrição da estrutura das informações através das marcações, que auxiliam a apresentação das informações como mostrado na Figura 2.

Figura 2 – Exemplo de informação marcada em HTML

```
[article pii=nd doctopic=oa language=es ccode=br1.1 status=1 version=3.1 type=tab order=04
seccode=RESP020 sponsor=nd stitle='Rev. Esp. Salud Publica' volid=74 issueno=4
dateiso=20000800 fpage=351 lpage=359 issn=1135-5727]

[front]
[titlegrp]

[title language=es] Utilización de anfotericina B no convencional en el Hospital Clínico de San
Carlos [/title]

[/titlegrp]
[authgrp]

[author role=nd rid='a01'] [fname] Concepción [/fname] [surname] Prieto Yerro [/surname] [/author]

[/authgrp] (2).
(1) [aff id='a01' orgname='Hospital Clínico de San Carlos' orgdiv1='Servicio de Farmacología
Clínica'] Servicio de Farmacología Clínica. Hospital Clínico de San Carlos. [city] Madrid [/city] . [/aff]

[abstract language=es] FUNDAMENTO: La anfotericina B es [...]

MÉTODOS: Se desarrollaron unas normas de uso de la anfotericina [...]

RESULTADOS: En el 54% de los tratamientos se hizo [...]

CONCLUSIONES: La evaluación retrospectiva ha mostrado que existe un elevado porcentaje de
tratamientos que no se adecuan a lo recomendado en las normas de prescripción. [...]

[/abstract]

Palabras clave: [keygrp scheme=nd] [keyword type=m language=es] Anfotericina
B [/keyword] [keyword type=m language=es] Infecciones fúngicas [/keyword] . [keyword type=m
language=es] Estudio de utilización de medicamentos [/keyword] . [/keygrp]
```

Fonte: Adaptado de
SciELO (sd) <http://old.scielo.org/php/level.php?lang=pt&component=56&item=23>

Segundo Alesso e Smith (2005, *apud* RAMALHO, 2017, p. 25), “a linguagem HTML tem como foco principal a apresentação dos documentos, e em suas primeiras versões não fornecia nenhum tipo de suporte para a representação semântica dos conteúdos”. Desde seu surgimento, algumas versões foram criadas para aperfeiçoar seu desempenho. O HTML+, HTML 2.0 e HTML 3.0, foram aprimorados com algumas mudanças para enriquecer as possibilidades desta linguagem.

Em 1997 o World Wide Web Consortium (W3C) publicou a versão HTML 3.2, definindo-a como uma recomendação oficial. Em 1999, o HTML foi atualizado para a versão 4.0, passando a incorporar recursos da eXtensible Markup Language (XML) em sua estrutura, dando origem ao eXtensible HyperText Markup Language (XHTML). Segundo Martins (2018. p. 31), esta pode ser considerada a grande mudança no HTML que viabilizaria a criação de páginas mais ricas estrutural e semanticamente.

Contudo, sua concepção baseada em marcações fixas das tags, ainda representava limitações para a representação semântica de informações, pois sendo uma linguagem desenvolvida principalmente para a apresentação de documentos, apenas define a forma como a informação é exibida, não se preocupando com o significado da informação.

De acordo com Almeida (2002), o HTML, ao servir à apresentação de páginas Web possibilita que os dados sejam visualizados em uma estrutura simples e de fácil leitura. Porém, sua característica estática limita a representação semântica de informações e consequentemente sua recuperação.

Num contexto cada vez mais voltado ao gerenciamento de dados através de diferentes aplicações que requerem formatos mais flexíveis, a tecnologia HTML não foi suficiente para suprir novas demandas por informações estruturadas.

6.3 eXtensible Markup Language – XML

Em 1994, World Wide Web Consortium (W3C) passou a trabalhar com objetivo de desenvolver protocolos para aperfeiçoar os diferentes erros e dificuldades que o HTML apresentava.

Em 1996, o (W3C) formou um comitê para resolver algumas limitações do HTML. Esse comitê identificou três aspectos do HTML que precisavam de atenção: Extensibilidade – que permitiria que os autores definissem suas próprias tags;

Estrutura – na qual os elementos pudessem ser aninhados dentro de outros elementos, como título dentro do artigo, capítulo ou citação, e Validação - permitindo que os autores verifiquem automaticamente a estrutura de um determinado arquivo contra uma estrutura padrão específica. A partir desses critérios, Jon Bosal entre os anos de 1966-1998 liderou um grupo de trabalho que desenvolveu eXtensible Markup Language – XML (KANE; MAKUCH, 1998).

O XML foi definido como um padrão de marcação para ser utilizado na Internet, constituindo-se uma versão simplificada da SGML, cuja característica principal é a possibilidade de definir e criar marcadores e atributos que vão para além de esquemas de marcação mais restritos como é o caso do SGML.

Tanto o HTML quanto o XML derivam do SGML, porém há diferenças significativas nas suas funções principais. Diferentemente das tags HTML, que tem como função principal a apresentação da informação em páginas Web, o XML é uma linguagem orientada para dados que descreve as entidades informacionais minuciosamente, através dos elementos e atributos.

O XML foi pensado para resolver as limitações da linguagem HTML, isto é, enquanto o HTML tem como objetivo controlar a forma com que os dados serão exibidos, o XML se concentra na descrição dos dados do ponto de vista semântico.

Marchal (2000 *apud* Bax, 2001), cita que a capacidade descritiva do XML possui vantagens significativas em relação ao HTML, ressaltando que sua aplicação pode ser útil na manutenção e gerenciamento de grandes sites, no intercâmbio da informação entre organizações e em dispositivos computacionais alternativos como os palmtops, handhelds etc.

Entralgo-Amaro e Aguiar-Cedeño (2009) apontam que a ausência de semântica em recursos informacionais como o HTML interfere na recuperação da informação:

Si en HTML hay que insertar metadatos para crear un mecanismo con el fin de recuperar un documento satisfactoriamente, en XML los metadatos ya vienen definidos desde su esencia. Lógicamente, la recuperación de información será más exacta sobre documentos donde las «etiquetas» están, única y exclusivamente, dirigidas a definir el contenido.

Segundo Almeida (2002) e Furgeri (2006), o armazenamento digital de documentos sem estrutura, transforma-se num estoque difícil de ser recuperado. Os autores abordam que os aspectos relevantes sobre a capacidade descritiva do XML, apontam para a possibilidade de inserção de conteúdo semântico que possibilitam melhorias significativas associadas a criação de metadados estruturados. Sendo estes, recursos de vital importância para a recuperação e intercâmbio de informação na web.

As considerações sobre o XML levam ao outro ponto de destaque, que vai além da facilidade de processamento e sua varredura por motores de busca na internet de forma mais eficiente, que é a interoperabilidade de informações.

O avanço das redes possibilitou o acesso à informação simultaneamente em diferentes tempos e contextos, ampliando desta forma seu alcance. A interoperabilidade é um componente primordial para a difusão da informação produzida de forma despesa entre diferentes instituições. Entende-se por interoperabilidade a capacidade de bases de dados compartilharem e trocarem informações a partir de diferentes plataformas de hardware e software.

Sobre este aspecto Furgeri (2006, p.233) destaca que:

Por meio da elaboração de estruturas padronizadas, torna-se perfeitamente possível que diversas comunidades consigam fazer intercâmbio e comparação de dados, além de otimizar a recuperação de informações.

Baseando-se em XML, diferentes instituições podem elaborar seus catálogos de forma padronizada e possibilitar que uma ferramenta de software desenvolvida para a Internet possa realizar buscas em acervos distintos de forma integrada.

A principal característica do XML é a possibilidade de agregar conteúdo semântico à informação, como mostra a Figura 3. Ao invés de priorizar apenas a apresentação a estruturação e agregação de conteúdo semântico ocorre através de “metas tags” que, por sua vez pode ter um conjunto de especificações denominados atributos, ou seja, componentes que refinam o valor semântico do dado marcado de maneira simplificada, que podem favorecer processos automatizados de recuperação e intercâmbio de dados.

Figura 3 - Informação estruturada em XML com o uso de atributos

```

<trans-abstract xml:lang="es">
  <title>RESUMEN</title>
    <sec>
      <title>FUNDAMENTO</title>
      <p>La anfotericina B es le tratamiento</p>
    </sec>
  <sec>
    <title>MÉTODOS:</title>
    <p>Se desarrollaron unas normas de uso de la[...]</p>
  </sec>
  <sec>
    <title>RESULTADOS:</title>
    <p>EL 54% de los tratamientos se hizo [...]</p>
  </sec>
  <sec>
    <title>CONCLUSIONES:</title>
    <p>La evaluación retrospectiva ha mostrado[...]</p>
  </sec>
</trans-abstract>

<kwd-group xml:lang="es">
  <title>Palabras clave: </title>
  <kwd>Anfotericina B</kwd>
  <kwd>Infecciones fúngicas</kwd>
  <kwd>Estudio de utilización de medicamentos</kwd>
  <kwd>Depresión</kwd>
  <kwd>Ansiedad</kwd>
</kwd-group>

```

Fonte: a autora

As linguagens de marcação são recursos não proprietários, isto significa que podem ser amplamente utilizadas e adaptadas para finalidades distintas no contexto de uma determinada instituição. A principal vantagem do XML que o difere das outras duas linguagens é que seu surgimento foi pensado para ser um mecanismo comum de codificação, transferência e armazenamento de metadados na internet e em sistemas automatizados de informação.

6.4 Document Type Definition - DTD

O Document Type Definition -DTD (Figura 4) especifica um conjunto de regras (sintaxe, estrutura e vocabulário) com as quais um documento deve estar em conformidade, sejam eles SGML, XML ou HTML. Isto é, a DTD define como as tags devem ser interpretadas, quais as regras que restringem o uso de cada uma delas na

estrutura de marcação e até mesmo a ordem em que as marcas devem aparecer no documento.

Figura 4 - Exemplo de especificações de um arquivo .DTD

```
<!-- ===== -->
<!-- MODULE:   Journal Publishing DTD           -->
<!-- VERSION:  ANSI/NISO JATS Version 1.0 (Z39.96-2012)  -->
<!-- DATE:    July 2013                          -->
<!--                                     -->
<!-- ===== -->

<!-- ===== -->
<!--          PUBLIC DOCUMENT TYPE DEFINITION          -->
<!--          TYPICAL INVOCATION                      -->
<!--                                     -->
"//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.1d1 20130915//EN"
  Delivered as file "JATS-journalpublishing1.dtd"
  Available at:
  http://jats.nlm.nih.gov/publishing/1.1d1/JATS-journalpublishing1.dtd
  -->
<!-- ===== -->

<!-- ===== -->
<!-- SYSTEM:   Journal Publishing DTD of the JATS DTD Suite  -->
<!--                                     -->
<!-- PURPOSE:  DTD for creation of new journal articles      -->
<!--                                     -->
<!--          The Journal Publishing DTD is an application of -->
<!--          the ANSI/NISO Z39.96 Journal Publishing Tag Set. -->
<!--          IT is a subset of the Journal Archiving and      -->
<!--          Interchange DTD that is optimized for the        -->
<!--          creation or publishing of existing journal       -->
<!--          articles and some non-article material such as    -->
<!--          product and book reviews in XML. It describes   -->
<!--          both the metadata for a journal article and       -->
<!--          the full content of the article.                  -->
<!--                                     -->
<!--          This DTD was constructed using the modules in the -->
<!--          JATS DTD Suite.                                   -->
```

Exemplo de DTD: <https://jats.nlm.nih.gov/publishing/1.0/dtd.html>

Um arquivo baseado em linguagem de marcação pode ser bem formado, porém “não válido” se estiver em desacordo com as definições da DTD. Essas definições são declaradas no elemento DOCTYPE. Ao serem submetidos ao um

analisador denominado parser, a lógica de validação considera toda a estruturação do artigo com base nesta informação (Figura 5)².

Figura 5 - Exemplo de DTD declarada na tag DOCTYPE de um arquivo XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.0
20120330//EN" "JATS-journalpublishing1.dtd">
```

Current Version (Standard):
NISO JATS Version 1.2: <https://jats.nlm.nih.gov/publishing/1.2/>

Previous Versions:

- NISO JATS Version 1.2d2 (Draft): <https://jats.nlm.nih.gov/publishing/1.2d2/>
- NISO JATS Version 1.2d1 (Draft): <https://jats.nlm.nih.gov/publishing/1.2d1/>
- NISO JATS Version 1.1 (2016 Standard): <https://jats.nlm.nih.gov/publishing/1.1/>
- NISO JATS Version 1.1d3 (Draft): <https://jats.nlm.nih.gov/publishing/1.1d3/>
- NISO JATS Version 1.1d2 (Draft): <https://jats.nlm.nih.gov/publishing/1.1d2/>
- NISO JATS Version 1.1d1 (Draft): <https://jats.nlm.nih.gov/publishing/1.1d1/>
- NISO JATS Version 1.0 (2012 Standard): <https://jats.nlm.nih.gov/publishing/1.0/>
- NISO JATS Version 0.4 (Draft): <https://jats.nlm.nih.gov/publishing/0.4/>

Adaptado de Panepucci (2019). Disponível em

<https://econtents.bc.unicamp.br/eventos/index.php/enapp/article/view/1564>

Embora não seja obrigatório que um documento XML tenha uma DTD declarada em DOCTYPE, arquivos que precisam ser validados necessitam que isso seja feito, como forma de garantir a conformidade estrutural do documento.

² O parser é um analisador sintático. Sua função é ler uma entrada de dados que possuem certas regras específicas a partir de uma determinada DTD que especifica a estrutura de um arquivo. Geralmente é utilizado para identificar erros estruturais e recusar as entradas que não estejam de acordo com as regras estabelecidas pela DTD.

6.5 As linguagens de marcação e os periódicos científicos

Segundo Lancaster (1995), com o avanço da internet, muitos periódicos ascendiam ao meio eletrônico como cópias escaneadas de suas versões impressas. Certamente, a publicação eletrônica de periódicos favoreceu a rápida publicação e resultados de pesquisa se comparado ao seu formato antecessor.

Como foi abordado no capítulo sobre comunicação científica, os periódicos desde o seu surgimento em 1665 até o atual contexto tecnológico passaram por profundas transformações. Com o tempo eles foram absorvendo novas tecnologias e novos formatos, mais adequados ao ambiente em rede, vislumbrando desta forma, possibilidades e características, tais como melhorias na apresentação em interfaces e inserção de hiperlinks.

Em determinado momento da evolução dos periódicos científicos, recursos computacionais mais sofisticados passaram a agregar o processo de produção e publicação, dentre eles estão a inserção das linguagens de marcação.

Algumas iniciativas voltadas ao desenvolvimento das linguagens de marcação que serão abordadas neste capítulo, concentraram seus esforços para incluir novas potencialidades à publicação eletrônica de periódicos.

6.5.1 AAP *Electronic Manuscript Standard* - AAP DTD

O interesse em desenvolver um padrão de marcação para periódicos com vistas a facilitar o intercâmbio de dados entre instituições remonta ao ano de 1983, quando a Association of American Publisher- AAP, uma organização representada por publicadores de livros e periódicos norte americano, estabeleceu o projeto *Electronic Manuscript Project*, que tinha como objetivo principal desenvolver um SGML padrão para o intercâmbio de dados de manuscritos.

A justificativa deste projeto era encorajar autores e comunidade envolvida nas publicações acadêmicas a usar o SGML na produção de manuscritos, uma vez que os processos editoriais de produção poderiam ser simplificados independente de outros sistemas eletrônicos se fossem baseados em formatos não proprietários. (CAVE, 2003).

O projeto atraiu a participação de grandes instituições, como a Library of Congress e American Society of Indexes. Como resultado, foi criado em 1985 o AAP

DTD que incluía padrões de marcação para monografias e publicações seriadas baseadas em SGML. A AAP DTD tornou-se em 1988, o padrão ANSI / NISO Z39.59 (Electronic Manuscript Preparation and Markup) do American National Standards Institute.

6.5.2 ISO 12083 - Serial DTD

A ISO 12083 Serial DTD foi desenvolvida pela Association of American Publishers (AAP) ao final dos anos 1980. Ela foi lançada como um padrão ANSI em 1988 e se tornou um padrão ISO em 1993. Sua última atualização ocorreu em 1995.

A ISO 12083 surgiu como um padrão internacional para intercâmbio de documentos entre autores e editores, fornecendo quatro definições de tipo de documento destinadas à criação e intercâmbio de livros, artigos e publicações seriadas em linguagem SGML. Ela foi projetada para apoiar publicações científicas, incluindo a marcação de resumos, autoria, afiliações, conteúdo textual, imagens e citações, que eram usados principalmente por editores técnicos para gerenciar o conteúdo em um nível apropriado de aplicação interna, assim como o reuso de dados; por exemplo, listas de autores e referências que podiam ser compiladas e usadas para verificação ou publicações subsequentes (MURRAY-RUST; RZEPA, 2003,).

Cada organização tem necessidades específicas que devem ser atendidas em uma DTD, porém, muitos dos editores que utilizaram a ISO 12083, mesmo que agregando algumas modificações, tiveram limitações para adotá-la definitivamente. Esta questão foi abordada em pesquisa conduzida pelo grupo de trabalho ISO 12083, de 1998, no qual constatou-se que os usuários tinham uma percepção de que essa DTD era muito complicada e pouco flexível para representar as informações necessárias de periódicos eletrônicos, principalmente as relacionadas à estruturação de dados essenciais para arquivamento (INERA, 2001).

No Brasil as linguagens de marcação em SGML estiveram presentes nos periódicos indexados na biblioteca eletrônica SciELO. Esta biblioteca surgiu em 1997, como resultado de um projeto de pesquisa da FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, em parceria com a BIREME - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. Segundo Packer (1998), os objetivos principais desse projeto visavam a disponibilização de artigos em

texto completo na internet, armazenamento, a avaliação da produção científica dos periódicos eletrônicos brasileiros que compunham a coleção.

De 1997 a 2012 a linguagem de marcação dos periódicos no SciELO utilizava a *DTD SciELO Article* que era baseada na norma ISO 12083. O uso desta DTD restringia-se à marcação dos elementos bibliográficos como título, autoria, resumos e palavras-chave e as referências bibliográficas de cada artigo (PACKER, 1998; 2014).

A partir de 2014, o SciELO passou a requerer o formato XML como requisito técnico para ingresso e permanência dos periódicos. Segundo Packer *et al* (2014) a metodologia de marcação anterior já não atendia às demandas atuais de estruturação, intercâmbio e apresentação dos textos dos artigos e outros tipos de documentos científicos. Essa mudança teve por objetivo otimizar seus processos técnicos internos relacionados à extração e reuso de metadados dos artigos que eram indexados na base.

6.5.3 Modular Application for journals - MAJOUR

O SGML orientado as publicações científicas serviu como base para a primeira DTD desenvolvida especificamente para marcação de periódicos científicos.

Essa DTD, denominada "*Modular Application for journals*" - MAJOUR foi publicada em 1991 como resultado dos trabalhos do *European Workgroup for SGML (EWS)*. Esse projeto envolvia publicadores como Elsevier, Science, Wolter Kluwer Academic Publishers e Springer Verlag que se basearam nos resultados dos trabalhos anteriormente desenvolvidos pela AAP (CAVE, 2003).

O desenvolvimento da MAJOUR se deu principalmente face à maior demanda por metadados, recuperação e intercâmbio de informação em diferentes bases de dados online, cada vez mais crescente entre os publicadores. Ainda segundo Cave (2003), a MAJOUR teria aportado a criação de diversas outras DTDs posteriores que foram criadas para aplicação em periódicos científicos.

6.5.4 Simplified SGML for Serial Headers - SSSH DTD

Outra DTD voltada ao objetivo de definir um padrão para a marcação de periódicos foi a *Simplified SGML for Serial Headers* SSSH DTD. A SSSH - SGML foi desenvolvida pelo grupo de trabalho *Organization for Article Standards In Science* - OASIS em nome da Book Industry Communications, órgão de padronização da

indústria de publicação de livros e publicações seriadas do Reino Unido e publicada em 1996.

No início da década de 1990, assim como o DTD (MAJOUR), a SSSH foi projetada para a marcação dos cabeçalhos de artigos científicos, fornecendo modelagem de informações bibliográficas sobre itens de publicações que eram frequentemente chamadas de modelos de “cabeçalho”. A diferença entre MAJOUR e SSSH era que esta possibilitava a marcação de afiliações institucionais. Ambos tiveram aceitação razoável por um tempo, porém não houve adesão prolongada ao modelo original por editores e publicadores de periódicos pois não atendiam de maneira abrangente às necessidades de marcação de metadados demandadas pelos editores (BROWN, 2003; CAVE, 2003).

Muitos grandes editores como Spring Verlag, Elsevier e Wiley passaram a desenvolver suas próprias DTDs em SGML se utilizando principalmente das recomendações da ISO 12083 DTD article e da MAJOUR, mas, agregando modificações e especificações próprias mais adequadas às suas necessidades (ROSENBLUM, 2015; WUSTEMAN, 2003b).

6.5.5 PMC Document Type Definition – PMC DTD

Quando do surgimento da base PubMed Central - PMC em 1999, o único requisito técnico para indexação era que os editores fornecessem os artigos em formato General Markup Language (SGML) ou Extensible Markup Language (XML), juntamente aos arquivos de imagem.

Os periódicos indexados eram baseados em DTDs variadas, o que ocasionava dificuldades técnicas de processamento e renderização das páginas pelos softwares utilizados para disponibilização dos artigos no site do PMC. Segundo Beck e USDIEN (2013), as dificuldades técnicas de compatibilidade associadas à multiplicidade de DTDs até então utilizadas pelos publicadores, representava um grande entrave técnico.

A partir disso, chegou-se conclusão de que um único modelo de estruturação de artigo poderia reduzir as dificuldades técnicas de processamento de arquivos na base, e foi este o impulso inicial para a criação de um padrão comum que fosse mais

facilmente gerido pelos sistemas de processamento de artigos publicados no Pubmed Central, dando origem ao *PMC Document Type Definition*.

A PMC Document Type Definition (DTD), também conhecida como pmc-1.dtd, foi uma DTD baseada na keton.dtd (SGML) e na article.dtd do Biomed central (xml). Ela foi elaborada com base nos dois principais modelos de arquivo que estavam sendo submetidos ao PMC na época e seu foco principal era a representação online dos artigos.

Em 2001, o Projeto Harvard University Library E-Journal Archiving Project encomendou um estudo sobre a viabilidade de ter um DTD que pudesse ser usado também para arquivar todos os periódicos eletrônicos. A questão que era colocada naquele momento estava na busca por uma mesma estrutura lógica e semântica para os textos científicos, que fosse para além da representação da informação no site e que atendesse também necessidades de arquivamento (ABRAMS, ROSENBLUM, 2003; BECK, 2011).

Iniciou-se então um trabalho de mapeamento das principais características das DTDs até então utilizadas por grandes editores de periódicos. A Inera, Inc., Belmont, Massachusetts, foi a instituição incumbida de preparar um relatório técnico sobre as DTDs baseando-se nos seguintes editores: Instituto Americano de Física, BioOne, Blackwell Science, Elsevier Science, Highwire Press, Institute of Electrical and Electronics Engineers, Nature Publishing Group, University of Chicago Press e John Wiley & Sons. Com base nessa análise, relatório concluiu que poderia ser criado um único DTD que pudesse acomodar qualquer artigo de periódico eletrônico para fins de arquivamento, pois nenhuma das DTDs analisadas no estudo atendeu a todos os requisitos (BECK; RANDALL, 2013; BECK; USDIEN 2013).

A partir de modificações em sua primeira versão, surgiu então a pmc-2.dtd, cujo objetivo era maior flexibilidade e a inserção de novos elementos e atributos descritivos em XML que pudessem dar conta mais precisamente do conteúdo textual e gráfico dos artigos científicos para fins de apresentação e arquivamento.

A principal mudança nesta segunda versão era que a DTD passaria a ser baseada apenas na linguagem XML e não mais em SGML, como o era sua antecessora. Esse aspecto se dava pela característica extensível e maior flexibilidade

para representar e arquivar qualquer artigo de periódicos baseado em formato XML (INERA, 2001).

6.5.6 Journal Archive and Interchange Tag Suite - NLM DTD

Em 2003, foi lançada a *NLM Archiving and Interchange Tag Suite* que incluía dois conjuntos de tags para a marcação de artigos científicos: o modelo de “arquivamento e intercâmbio” e o modelo de “publicação” de periódicos. O primeiro foi planejado para marcar conteúdos já existentes, e o segundo mais prescritivo, foi projetado para ser utilizado para a criação novos conteúdos do artigo marcados apenas no formato XML.

Em 2005, foi lançada a Tag Suite versão 2.1 que contava com a inserção de um novo modelo de artigo: o modelo *Article Authoring*. Posteriormente, a versão 2.3 surgiu para compatibilizar os modelos até então elaborados. Simultaneamente com o lançamento desta, tomou-se a decisão de que a próxima atualização, incorporaria todas as alterações realizadas nas versões anteriores.

Em 2008 foi lançada a versão 3.0, uma tag suíte totalmente nova. Por ter agregado muitas modificações, ela não era compatível com suas versões anteriores. Essa nova versão, foi importante pois serviu como base para a criação da *Jornal Article tag suíte*, que será abordada no Capítulo 7.

Segundo Rosenblum e Golfman (2002), embora o XML tenha se tornado um padrão recomendado pelo (W3C) em 1998, boa parte dos principais publicadores ainda utilizavam DTDs próprias baseadas em SGML, até pelo menos 2002 na nas publicações científicas.

O aumento da informação científica produzida em ambiente digital alinhado a novas demandas por metadados estruturados contribuiu para mudanças significativas nas tecnologias que eram até então utilizadas na publicação de periódicos. As linguagens de marcação certamente tiveram papel importante nesse processo, o que levou as publicações a voltarem seu olhar em especial para o XML inclui diversas razões.

Kane e Makuch (1998), apontam que o amplo uso de SGML por profissionais da informação e editores foi limitado pela complexidade na adequação das DTDs às necessidades ou projetos específicos, além de altos custos envolvidos para colocar

em prática um sistemas baseados em SGML. Além disso, de acordo com Beck e Randall (2013) e Rosenblum (2015), há mais oferta de softwares e aplicações disponíveis criados para processar arquivos XML, se comparado com o SGML.

Segundo Rosenblum e Golfman (2002), o uso do XML em periódicos eletrônicos mostrava novas possibilidades de preservar a longo prazo as informações contidas nos periódicos eletrônicos. Valendo-se das características extensíveis e estruturais conferidas a este padrão, seria possível melhorias nos processos de arquivamento dos periódicos.

Para Brown (2003) o XML se configura por excelência na atividade de intercâmbio científico, que permite aumentar substancialmente a visibilidade e o alcance dos periódicos publicados na web.

Packer *et al*, (2014), aponta que a motivação para adotar o XML nas publicações científicas se baseou na busca por um formato que possibilitasse gerar versões tanto impressas quanto online, assim como a maior facilidade de intercambiar conteúdo.

A inserção das linguagens de marcação tiveram importante papel e serviram a automatização e melhorias significativas nos processos editoriais. O SGML permitiu a estruturação da informação de modo que pudessem ser mais facilmente geridos por recursos computacionais, mas sua complexidade limitou sua ampla adesão para uso generalizado. Por outro lado, o HTML não oferecia estabilidade e um nível de estruturação eficaz para documentos em grandes coleções de dados, mostrando limitações tanto para o arquivamento adequado de informações estruturadas quanto para a gestão de grandes volumes de informação. Mostrando-se com uma alternativa necessária que combinou o rigor de um com a simplicidade do outro, o uso de XML na publicação de periódicos eletrônicos emergiu como um ponto importante para as bibliotecas em função facilidade de manuseio e grande utilidade para geração de metadados estruturados que favorecessem os processos de armazenamento e gestão das informações (WUSTEMAN, 2003a).

7. ANSI/NISO Z39.96-2012 - JOURNAL ARTICLE TAG SUITE – JATS

O ANSI / NISO Z39.96-2012, também conhecido por Journal Article Tag Suite (JATS) é um padrão que descreve um conjunto de elementos e atributos usados para construir modelos XML de artigos de periódicos para arquivamento, publicação e autoria.

A JATS se tornou um Padrão Nacional Americano (ANSI / NISO Z39.96-2012) em agosto de 2012, e tem sua origem nas especificações da NLM DTD versão 3.0. Após a conclusão do grupo de trabalho *NLM Archiving and Interchange Tag Suite Working Group* que desenvolveu a versão 3.0 da NLM DTD lançada em novembro de 2008, a aderência a este padrão foi tão expressiva entre publicadores e editores que se iniciou então a criação em 2009 de um novo grupo com envolvimento da NISO, denominado *NISO Standardized Markup for Journal Articles Working Group*. Esse grupo foi responsável pela criação da primeira versão pré-NISO, que em essência era uma pequena atualização da NLM versão 3.0 (BECK; USDIN, 2013).

Em 2012, sob a padronização da ANSI/NISO Z39.96-2012, foi lançada oficialmente a *Journal Article Tag Suite*. Este padrão define elementos e atributos para descrever metadados do conteúdo completo de artigos de periódicos científicos e desde então vem sendo utilizado para agregar novos elementos e estruturas à marcação, com base nas necessidades reais de seus usuários.

Os conjuntos de tags JATS são mantidos por um comitê NISO denominado *JATS Standing Committee*. Sugestões de mudanças na JATS podem ser feitas por qualquer pessoa preenchendo um formulário aberto no site da NISO. O Comitê se reúne periodicamente para revisar as sugestões e algumas resultam em mudanças em um ou mais conjuntos de tags JATS ou na sua documentação completa.

Atualmente a JATS está em uso em mais de 25 países, incluindo: Austrália, Bélgica, Brasil, Bulgária, Canadá, Chile, Egito, Finlândia, França, Alemanha, Itália, Japão, Noruega, Rússia, Coreia do Sul, Estados Unidos entre outros (LAPEYRE, 2018).

Nas etapas iniciais de sua criação, centenas de periódicos foram examinados quanto à estrutura de metadados em mais de 45 modelos de periódicos XML e SGML. Os desenvolvedores da JATS observaram que 80% dos periódicos marcavam os

dados de maneira semelhante, e isso foi fundamental para a criação das tags sets. (LAPEYRE, 2018).

Acerca das vantagens sobre adoção da JATS, Beck e Randall (2013) destacam que a estruturação de artigos de periódicos a partir de um modelo já definido, facilita o intercâmbio entre editores, autores, conversão de dados e intercambio entre agregadores de conteúdo científico. Por estar disponível gratuitamente, permite que editores de pequenos periódicos possam criar artigos em XML significativamente mais rápido. Além disso a Journal article tag suite foi projetada para a criação de novos esquemas, sendo possível sua adequação para elaboração de conjuntos de metadados estruturados para diferentes projetos e objetivos.

Para Huh (2014) dentre as possibilidades que se surgem com o uso da JATS está a maior visibilidade dos periódicos, pontuando que artigos produzidos em XML-JATS podem ser incluídos em grandes em bases de dados internacionais de texto completo, como PMC ou ScienceCentral. Ainda, uma vez incluídos nessas bases, tornam-se mais facilmente rastreados pelo Google Scholar o que faz aumentar sua exposição a um número maior de pesquisadores. O autor também destaca que arquivos JATS são de mais fácil conversão para outros visualizadores e outros modelos de XML.

Cabe ressaltar, que a aceitação de periódicos nessas bases não se limita apenas aos aspectos técnicos, elas dependem também da avaliação sob aspectos pertinentes à qualidade dos conteúdos científicos. Investigando os requisitos técnicos para indexação nas bases SciELO, e PMC, identificou-se que apenas estruturar artigos em XML-JATS não garante o ingresso de periódicos nessas bases. Os arquivos devem estar criteriosamente estruturados sob a DTD específica da base no qual pretende a indexação para se atender aos requisitos técnicos de indexação.

7.1 Características das Tag sets - JATS

A JATS é composta por três conjuntos de tags. Esses conjuntos são chamados “Tag sets”, e agrupam elementos e atributos que especificam suas relações no contexto de marcação de um artigo, tais como a ordem em que devem aparecer e a cardinalidade dos elementos.

O ponto mais crucial da documentação dos elementos e atributos é entender quais deles são possíveis na estrutura que se pretende marcar, quantas vezes podem ocorrer e em qual ordem devem aparecer na construção de um artigo.

Cada uma das tag sets possui uma respectiva documentação chamada *Tag Libraries*, Figuras 6 e 7, na qual se descreve todos os elementos e atributos que são utilizados para a marcação em XML de artigos, assim como sua função.

Figura 6 – Página inicial da documentação da Journal Article Tag Suite



Tag set	Elementos	Atributos
Archiving y Interchange	254	135
Journal Publishing	256	134
Article Authoring	197	123

Fonte: a autora

Figura 7 - Tag library do conjunto Journal Publishing



Adaptado de <https://jats.nlm.nih.gov/publishing/tag-library/1.3d1/>

O mapeamento de elementos das três tag sets mostra que a maior parte deles é comum nas três versões, porém observa-se que Article Authoring Tag Set apresenta-se como a mais prescritiva em relação à totalidade de elementos das outras duas tag sets. A Tabela 1 mostra a compatibilidade de elementos, entre as três versões.

Tabela 1 - Compatibilidade entre elementos da três tag sets

Archi	Publ	Autho	Elementos
x	x		<abbrev-journal-title>
x	x		<alt-title>
x	x		<article-categories>
x	x		<article-id>
x	x		<article-version-alternatives>
x	x		<author-notes>
x	x		<compound-subject>
x	x		<compound-subject-part>
x	x		<corresp>
x	x		<count>
x	x		<counts>
x	x		<custom-meta>
x	x		<custom-meta-group>
x	x		<equation-count>
x	x		<event>
x	x		<event-desc>
x	x		<fig-count>
x	x		<fig-group>
x	x		<floats-group>
x	x		<front-stub>
x	x		<history>
x	x		<issue-sponsor>
x	x		<journal-meta>
x	x		<journal-subtitle>
x	x		<journal-title>
x	x		<journal-title-group>
x	x		<meta-name>
x	x		<meta-value>
x	x		<milestone-end>
x	x		<milestone-start>

Archi	Publ	Autho	Elementos
x	x		<note>
x	x		<notes>
x			<overline-end>
x			<overline-start>
x	x		<pub-date>
x	x		<pub-date-not-available>
x	x		<pub-history>
x	x		<publisher>
x	x		<ref-count>
x	x		<response>
x	x		<rp>
x	x		<sec-meta>
x	x		<series-text>
x	x		<series-title>
x	x		<sig>
x	x		<sig-block>
x			<string-conf>
x	x		<string-date>
x	x		<sub-article>
x	x		<subj-group>
x	x		<table-count>
x	x		<table-wrap-group>
x	x		<tex-math>
x	x		<trans-abstract>
x	x		<trans-subtitle>
x	x		<trans-title-group>
x			<underline-end>
x	x		<underline-start>
			<unstructured-kwd-group>
x	x		<volume-issue-group>

Na documentação de cada *tag set* são disponibilizados diagramas que retratam a estrutura geral dos artigos e orientam a marcação para cada finalidade, seja ela publicação, arquivamento ou autoridade. A simbologia que acompanha cada elemento no diagrama tem a função de especificar sua ocorrência numa determinada estrutura de artigo em XML.

?	Elemento opcional, zero ou uma ocorrência
+	Elemento obrigatório, pode ser repetir uma ou mais ocorrências
*	Elemento opcional pode ocorrer zero ou mais ocorrências
~	Elemento possui atributo

7.1.1 Journal Archive e Interchange Tag Set

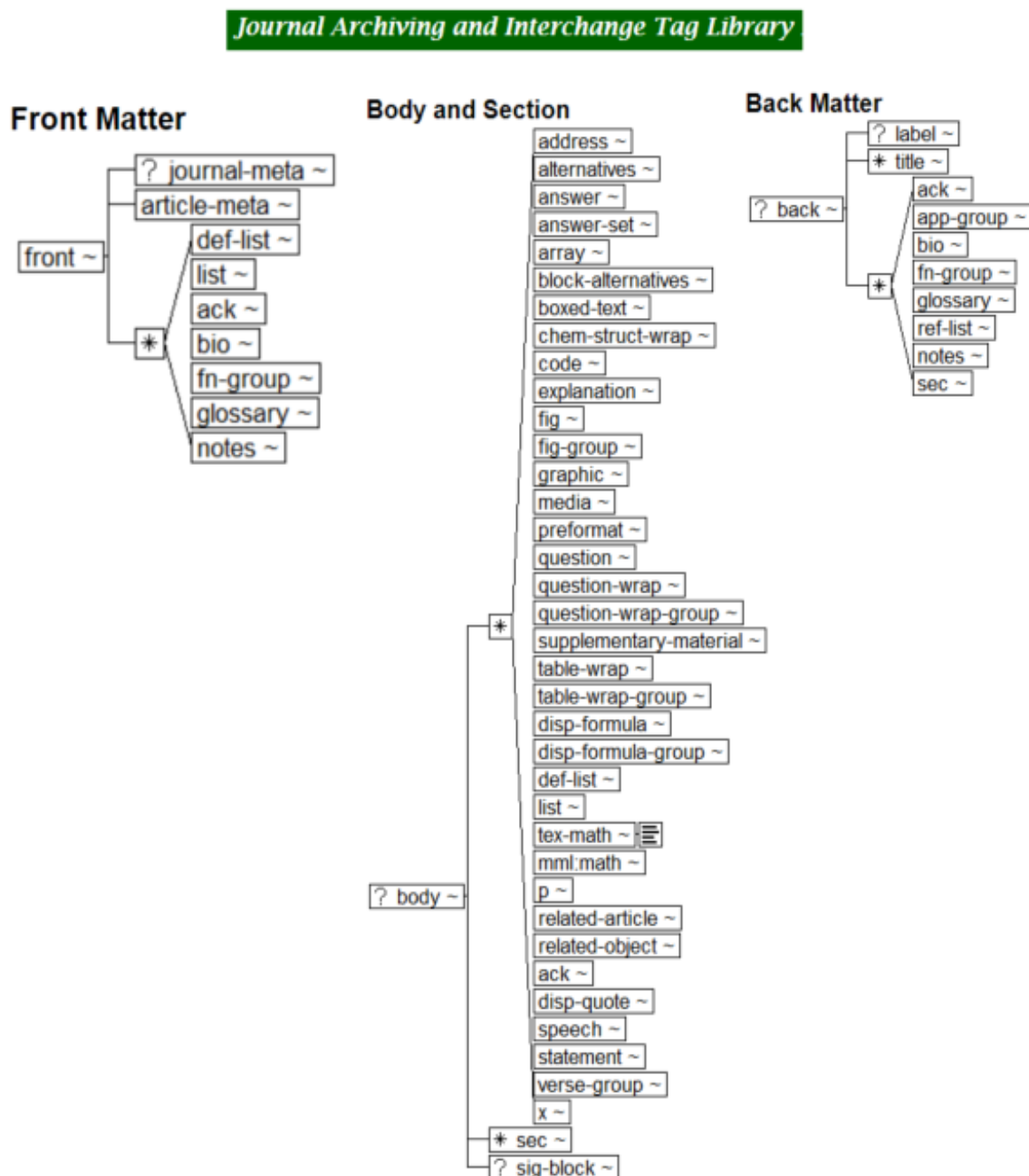
O *Journal Archive and Interchange Tag Set* fornece um conjunto de *tags* no qual os editores podem submeter o conteúdo marcado para a finalidade de arquivamento. Seus elementos foram expressamente criados para que as informações marcadas não sejam perdidas quando ocorre a conversão do artigo para outros modelos de XML (Figura 8). Um dos seus principais objetivos é preservar o conteúdo intelectual dos artigos de revistas, independentemente da forma como o conteúdo foi originalmente criado.

O modelo de arquivamento fornece um elemento para registrar o processamento que pode ter sido feito no XML pré-JATS (<x>), que pode ser útil para reter informações de documentos XML não JATS. O arquivamento é destinado a bibliotecas e arquivos que aceitam a entrada de uma ampla variedade de fontes e as converte JATS.

Este é o conjunto para marcação mais permissivo dentre as três versões da TAG suite. Sua principal característica é que ela permite que num processo de arquivamento automatizado, sejam capturados componentes semânticos e estruturais sem que haja necessidade de especificar uma sequência ou formato textual particular.

Além de preservar os metadados e o conteúdo do artigo, este conjunto de tags permite a preservação do conteúdo de informações como legendas, números, símbolos e pontuação nas referências bibliográficas.

Figura 8 - Diagrama da estrutura básica da Archiving and Interchange Tag Set



Fonte: Adaptado de <https://jats.nlm.nih.gov/archiving/tag-library/1.3d1/chapter/nfd-front-matter.html>

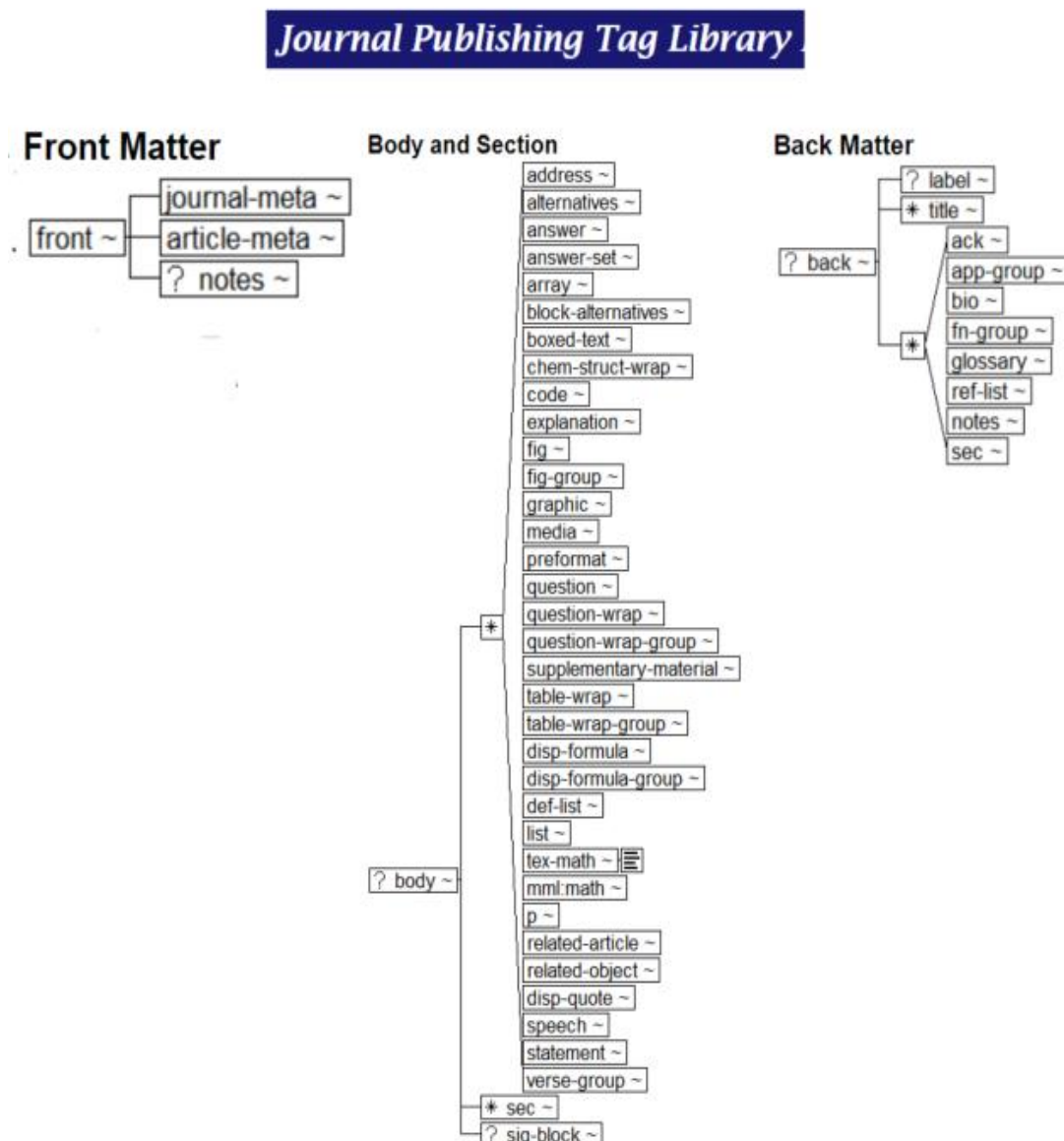
7.1.2 Journal Publishing Tag Set

O Journal Publishing Tag Set define os elementos e atributos que descrevem o conteúdo e metadados de artigos de periódicos em texto completo. Esse conjunto foi desenvolvido com base Journal Archive and Interchange Tag Set, mas com a diferença que foi orientado a ser um formato comum a publicação de artigos científicos (Figura 9). É um conjunto moderadamente prescritivo pois define mais sequências de elementos estruturados do que o conjunto voltado ao arquivamento.

Dentre suas possibilidades de uso desta versão está sua aplicação a grandes volumes informacionais que precisam ser normalizados ou sejam passíveis de processos de controle de conteúdo. A criação de arquivos neste formato pode ocorrer a partir de artigos que já foram publicados em outros formatos como PDF ou Word.

O conjunto de tags de publicação é uma boa alternativa para editores que convertem conteúdo de material de origem não XML, como formatos de composição proprietários e arquivos de processamento de texto. Ele permite a marcação de metadados que são criados durante o processo de publicação, como data de publicação, periódico, volume, edição e números de página, e é flexível o suficiente para acomodar a maioria dos estilos de editor.

Figura 9 - Diagrama da estrutura básica da Archiving Interchange Tag Set



Fonte: adaptado de <https://jats.nlm.nih.gov/publishing/tag-library/1.3d1/chapter/nfd-front-matter.html>

7.1.3 Article Authoring Tag Set

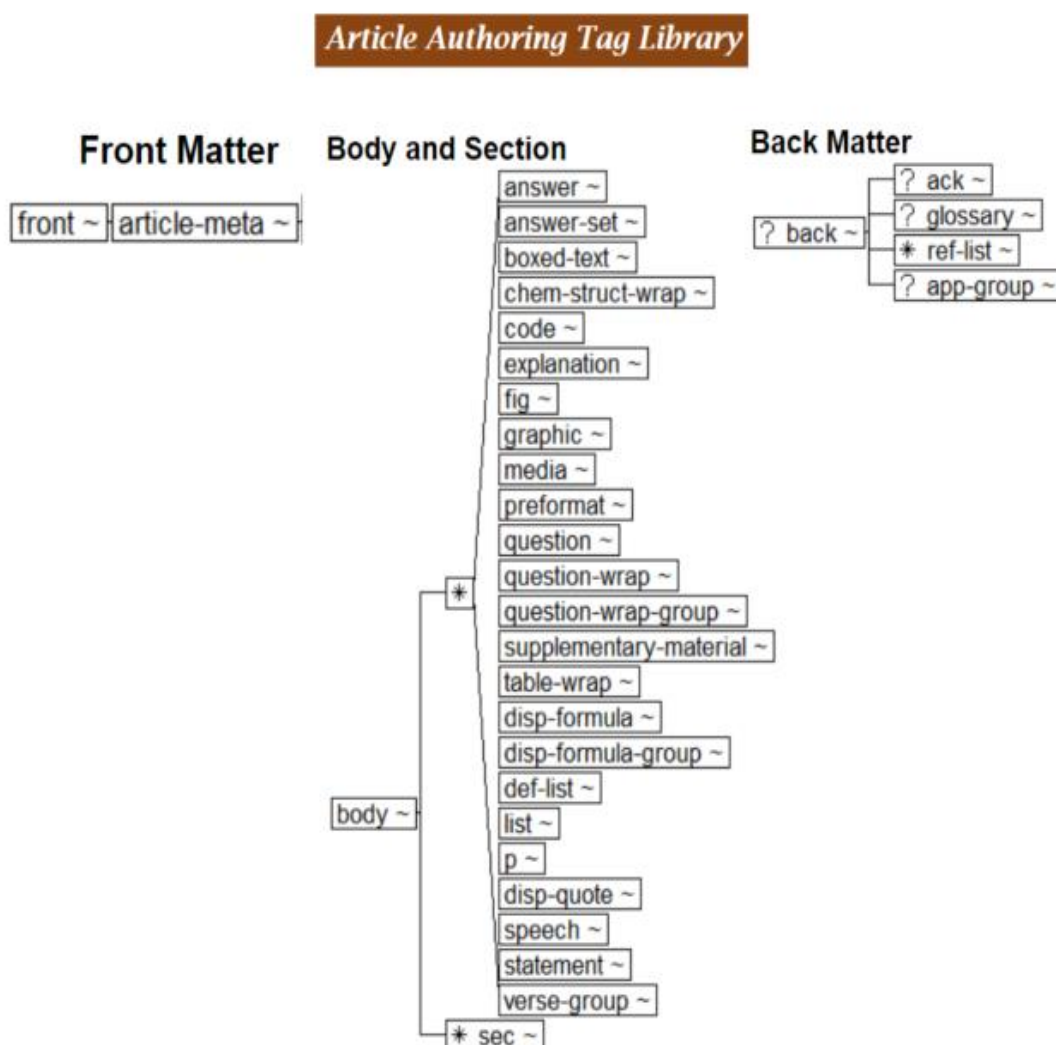
O conjunto de tags Article Authoring Tag Set define elementos e atributos que descrevem o conteúdo e os metadados de artigos focado nos dados de autoria.

Esse conjunto é mais prescritivo dos três modelos, porém além dos metadados do cabeçalho, permite a marcação de alguns elementos textuais e pós textuais para a descrição do conteúdo completo do artigo (Figura 10). Durante sua aplicação, o controle do conteúdo e a sequência dos itens devem ser rigorosamente observados,

isto significa que em sua definição, os elementos de marcação devem seguir em uma ordem específica limitando assim as opções para variações na estruturação de marcação dos artigos.

Esse conjunto não fornece marcação para metadados sobre o periódico, tais como data de publicação, número da página, histórico ou nome do periódico. Por ser um modelo cuja aplicação está focada no processo inicial de criação do artigo por seus autores, não contém elementos sobre o fascículo, como número, volume, ISSN ou nome do periódico.

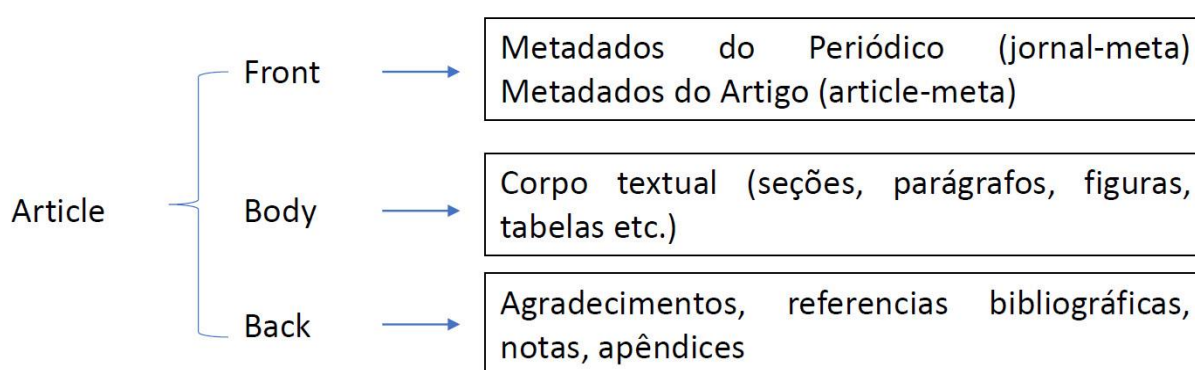
Figura 10 - Diagrama da estrutura básica da Article Authoring Tag Set



7.2 Análise de um artigo em formato XML-JATS

Um artigo científico estruturado em XML consiste basicamente em FRONT, BODY e BACK, conforme mostra a Figura 11. Como a marcação não pode ser feita de modo aleatório, cada uma dessas partes possui tags de marcação com elementos e atributos específicos descritos na documentação das *tag sets*.

Figura 11 - Estrutura geral de artigo científico



Fonte: a autora

O elemento FRONT descreve as informações do cabeçalho do artigo, que geralmente é o conteúdo da primeira página em sua versão PDF. O segundo, BODY descreve o corpo principal, que inclui conteúdos textuais e de imagem e o BACK descreve informações tais como referências bibliográficas, agradecimentos e dados de financiamento.

As *tags* de cada um desses conjuntos podem receber atributos. Como posto no capítulo sobre linguagens de marcação, a função dos atributos em uma estrutura XML é de refinar o conteúdo semântico do que está sendo marcado e possibilitar que esse conteúdo seja mais facilmente identificado por sistemas de processamento de informação automatizados.

FRONT

Um exemplo de informação que compõe o FRONT de um artigo pode ser os dados sobre os periódicos, título e autoria, resumo e palavras chaves dos artigos.

A Figura 12 representa uma estrutura sobre os elementos possíveis no conjunto “*jornal-meta*” definidos na tag suíte da JATS e sua respectiva replicação em formato XML. Essa figura demonstra a possibilidade de estruturação da informação a partir dos modelos de publicação e arquivamento, uma vez que esses dados não são passíveis de marcação no modelo de controle de autoria.

Figura 12 - Dados sobre o periódico estruturados em XML-JATS

```
<journal-meta>
  <journal-id journal-id-type="nlm-ta">J Venom Anim Toxins </journal-id>
  <journal-id journal-id-type="publisher-id">jvatitd</journal-id>

  <journal-title-group>
    <journal-title>Journal of Venomous Animals </journal-title>
    <abbrev-journal-title abbrev-type="publisher">J. Venom. Anim. </abbrev-journal-
      title>
  </journal-title-group>
  <issn pub-type="epub">1678-9XXX</issn>

  <publisher>
    <publisher-name>Centro de Estudos de Venenos Peçonhentos</publisher-name>
  </publisher>
</journal-meta>
```

Fonte: a autora

As *tags* de marcação contam com um conjunto de atributos específicos que refinam a natureza do dado marcado. A Figura 13 demonstra dados de autoria estruturados em XML-JATS com especificações de nome, sobrenome, função e contexto no qual se deu a autoria.

Figura 13 - Exemplo de dados de autoria estruturado em JATS

```

<contrib-group content-type="conference-editors">
  <contrib contrib-type="author">
    <name name-style="editor">
      <surname>Herrera</surname>
      <given-names>Gerardo</given-names>
    </name>
    <role>Conference Editor</role>
  </contrib>
</contrib-group>

```

Fonte: a autora

BODY

O BODY de um artigo científico, corresponde ao corpo textual do artigo. Ele pode ser composto por inúmeras informações, desde parágrafos, até informações mais complexas como tabelas e listas. (Figura 14 e 15)

Figura 14 - Parágrafo de artigo marcado em XML

```

<sec sec-type="discussion">
  <title>Discussion</title>
  <p>Upon examination of the reported advantages and
  disadvantages of JIF, a substantial difference was found in both the
  diversity and frequency of reporting. There were a significantly greater
  number of disadvantages recorded in comparison to advantages (18 vs. 9)
</p>
<sec>

```

Fonte: a autora

No caso quando houver tabelas, listas e quadros, a aplicação de tags em BODY tende a ser mais minuciosa, pois existem elementos descritivos para cada linha ou célula das tabelas que são codificadas em XML-JATS.

Figura 15 - Exemplo de tabela marcada em XML-JATS

```

<table rules="groups" frame="hsides">
  <tbody><tr>
    <td valign="top" align="left"></td>
    <td valign="top" align="left"></td>
    <td valign="top" align="left">Non-mix-OSA</td>
    <td valign="top" align="left"></td>
    <td valign="top" align="left"></td>
    <td valign="top" align="left">Mix-OSA</td>
    <td valign="top" align="left"></td>
  </tr>
  <tr>
    <td valign="top" align="left"></td>
    <td valign="top" align="left">Supine</td>
    <td valign="top" align="left">Non-supine</td>
    <td valign="top" align="left">p value</td>
    <td valign="top" align="left">Supine</td>
    <td valign="top" align="left">Non-supine</td>
    <td valign="top" align="left">p value</td>
  </tr>
  <tr>
    <td valign="top" align="left">OAI (/h)</td>
    <td valign="top" align="left">21.0&#x00B1;20.3</td>
    <td valign="top" align="left">5.7&#x00B1;11.9</td>
    <td valign="top" align="left">&lt;0.001</td>
    <td valign="top" align="left">30.4&#x00B1;15.5</td>
    <td valign="top" align="left">15.6&#x00B1;18.3</td>
    <td valign="top" align="left">&lt;0.001</td>
  </tr>
  <tr>
    <td valign="top" align="left">MCAI (/h)</td>
    <td valign="top" align="left">1.9&#x00B1;3.1</td>
    <td valign="top" align="left">0.6&#x00B1;2.3</td>
    <td valign="top" align="left">&lt;0.001</td>
  </tr>
</tbody>
</table>

```

Fonte: a autora

BACK

Em Back encontram-se principalmente referências bibliográficas, agradecimentos e materiais suplementares.

A Figura 16 mostra como essas informações podem ser exaustivamente detalhadas pois geralmente visam a atender demandas por metadados bastante específicas. Tais como a geração de índices bibliométricos ou estratégias de mineração de texto.

Figura 16 - Referência bibliográfica estruturada em XML-JATS

```
<ref id="B1">
<mixed-citation>Stojanovska J. The Figley Fellowship: An Entrance to Fundamentals.
Am J Roentgenol. 2015;204(4):2013-5. </mixed-citation>
<element-citation publication-type="journal">
  <person-group person-group-type="author">
    <name>
      <surname>Stojanovska</surname>
      <given-names>J</given-names>
    </name>
  </person-group>
  <article-title>The Figley Fellowship: An Entrance to Fundamentals</article-title>
  <source>Am J Roentgenol</source>
  <year>2015</year>
  <volume>204</volume>
  <issue>4</issue>
  <fpage>2013</fpage>
  <lpage>2015</lpage>
</element-citation>
</ref>
```

Fonte: a autora

Informações sobre agradecimentos podem conter além de texto, identificadores de bolsas, auxílio financeiros e instituições que subsidiaram a pesquisa. Neste caso é necessária uma marcação auxiliar que isola esses dados na estruturação do XML e os transfere para a estrutura de metadados sobre o artigo denominado <article-meta>, como mostrado na Figura 17.

Figura 17 - Marcação de dados de financiamento

```

<article-meta>
  <funding-group>
    <award-group>
      <funding-source>Fundação de Amparo à Pesquisa </funding-source>
      <award-id>001</award-id>
    </award-group>
    <funding-statement>Fundação de Amparo à Pesquisa do Distrito Federal, código
    de financiamento 001; concessão de bolsa depós-doutorado. Edital n 03/2016
    </funding-statement>
  </funding-group>
</article-meta>

```

Fonte: a autora

Os elementos e atributos nas Tag Library da JATS possibilitam a marcação do conteúdo inteiro dos artigos, mantendo as especificidades de cada conjunto de informação e o alto nível de detalhamento dos metadados.

Muitos arquivos públicos e privados aceitam ou requerem XML-JATS, incluindo: PubMed Central, Biblioteca Nacional Britânica, Biblioteca Nacional Australiana, Biblioteca do Congresso dos EUA, Portico, JSTOR entre outros (LAPEYRE, 2018).

O envio de arquivos a essas bases normalmente ocorre via protocolo FTP, mas antes de ser realizado é necessário que os arquivos estejam criteriosamente válidos e de acordo com as respectivas DTD.

Dada a complexidade que caracteriza os artigos científicos, como existência de textos longos com tabelas, equações e símbolos, a tarefa de verificação da consistência do xml por humanos seria praticamente inviável. Nesse sentido, a verificação do conteúdo marcados em XML-JATS pode ser realizada através de analisadores sintáticos de XML os chamados de parser que percorrem o documento e verificam se sua estrutura está de acordo com as regras de uma determinada DTD (DEITEL, 2008). Algumas bases que requerem este padrão costumam oferecer ferramentas de validação de arquivos XML-JATS. Foram identificados parsers em

acesso aberto desenvolvidos e disponibilizados para a validação de arquivos em formato XML-JATS nas bases SciELO, PMC³.

7.3 Como o padrão XML-JATS está sendo utilizado

Qualquer um dos conjuntos de tags da JATS pode ser expandido para atender às necessidades de projetos específicos. Isso significa a possibilidade de criar regras próprias a partir dos conjuntos de tags e atributos existentes na JATS.

Intuições como bibliotecas eletrônicas e agregadores de periódicos vem utilizando a JATS em seus processos editoriais para otimizar processos relacionados a publicação de artigos científicos que são indexados.

Esse capítulo se dedica a investigar casos de usos de algumas instituições que vem agregando esse padrão em seus processos de publicação e tratamento da informação.

SciELO

SciELO - Scientific Electronic Library Online é uma base de periódicos científicos brasileiros criada em 1997 a partir de um projeto de pesquisa da FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, em parceria com a BIREME - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. Dentre seus objetivos está maior visibilidade, armazenamento, disseminação e avaliação da produção científica brasileira em acesso aberto.

A partir de 2014 o SciELO passou a adotar a estruturação dos textos completos em formato XML-JATS. Com base neste padrão, a instituição desenvolveu uma

³ SciELO Package Maker: Plataforma Windows e tem como função realizar a validação de conjuntos arquivos com base na DTD JATS e especificações SciELO PS. Permite a visualização de tabelas, equações e figuras para revisão em formato HTML e gera um relatório de Erros e Avisos sobre possíveis ajustes e correções.

PMC Article Preview: Plataforma Online que realizada a validação em arquivos individuais do DTD JATS, e especificações PMC Guidelines. Permite a visualização HTML do texto completo para revisão.

metodologia própria de marcação denominada *SciELO Publishing Schema*, a fim de atender demandas específicas de normalização e reuso de dados.⁴

Esse Schema é composto por regras que especializam aspectos da marcação XML-JATS Publishing. Desde sua criação, ele vem sendo aprimorado com a inclusão de novas tag e estruturas e atualmente está em sua Versão 1.10 de março de 2020, que conta com uma lista de 125 elementos baseada na Tag set Publishing da JATS (SciELO, 2019).

Com o XML-JATS associado às definições pré-estabelecidas no schema SciELO, é possível otimizar o reuso dos metadados específicos. O SciELO estabeleceu nas especificidades deste schema, o detalhamento na marcação de afiliações institucionais, dados de financiamento e referências bibliográficas dos artigos. Isto é requerido para o acompanhamento da origem institucional e geográfica das pesquisas, necessário para geração de indicadores de produção científica por instituição e suas unidades principais e a criação de base de dados bibliométricos de citação nas publicações indexadas, assim como a identificação das fontes de financiamento declaradas nas pesquisas que são publicadas na base (SANTOS, 2015; SciELO, 2017).

Europe PMC

Europe PMC é uma das principais bases de dados de literatura científica sobre ciências da vida e conta com 3,0 milhões de artigos de periódicos em texto completo marcados em XML-JATS. Dentre seus objetivos, está o engajamento em apoiar a descoberta e reutilização de dados de pesquisa expressos nos artigos.

Segundo Parkin (2019), as agências de financiamento precisam identificar e rastrear os resultados da pesquisa para avaliar o impacto de seu financiamento entre diferentes disciplinas e áreas geográficas. Nas ciências da vida, isso é fator principal da avaliação da pesquisa e, portanto, a identificação de dados de financiamento é de vital importância.

⁴ Os esquemas XML permitem restrições mais precisas e permitem a definição de tipos de dados mais específicos.

Para tal, elaboraram um trabalho de análise sobre tendências de compartilhamento de dados expressas nos artigos, considerando aspectos sobre as condições e licenças de uso e onde os dados que sustentam os artigos são mais frequentemente hospedados. Inicialmente foi aplicado um algoritmo para rastrear as ocorrências de palavras como Data Availability, Data Accessibility, Data, Database survey e identificaram que esses dados ocorriam em partes variadas nos textos dos artigos o que dificultava a elaboração de estratégias de descoberta de dados.

Como forma de otimizar esse processo, foi formalizado desde 2018 a marcação dessas informações com as tags <sec> contendo o atributo @ sec-type = “data-availability”. Parkin (2019) frisa que o formato padrão fornecido pelo JATS foi primordial nesta atividade, pois fornece elementos descritivos específicos para a marcação de tais informações de modo que elas sejam mais facilmente filtradas através de estratégias de mineração de texto.

J-STAGE

J-STAGE é uma das principais plataformas de periódicos eletrônicos de informações científicas e técnicas no Japão, desenvolvida e gerenciada pela Japan Science and Technology Agency (JST). O objetivo do J-STAGE é acelerar a internacionalização da difusão e circulação de informações científicas e tecnológicas japonesas, através de um fluxo que abrange da submissão à divulgação de informações científicas e tecnológicas na Internet.

Um grande aprimoramento do J-STAGE é que ele adotou totalmente o JATS e passou a requerer que todos os periódicos incluam os metadados e as informações bibliográficas neste formato durante o processo de submissão de artigos.

Katoh *et al* (2012) considera que a adoção do formato XML para periódicos no Japão ocorre lentamente devido aos altos custos e dificuldades técnicas para produzir manuscritos nesse formato. A fim de minimizar o impacto dessas questões sobre as instituições que fornecem as informações ou empresas prestadoras que fazem o serviço de criação de metadados, o J-STAGE passou desde 2012, a oferecer à sociedade acadêmica do Japão um serviço web para auxiliar a automatização do registro de informações bibliográficas dos artigos que geram metadados em formato XML- JATS.

O formato de entrada para esta ferramenta é o PDF do artigo. Este PDF é analisado e os metadados das informações bibliográficas que abarcam título, autores, afiliações, resumo e referências de um artigo, além de informações referentes ao periódico, incluindo o título o, ISSN, data de edição etc., são extraídos em formato XML -JATS.

A ferramenta também possui um recurso de edição de XML. Com este recurso o formato de saída pode ser verificado e editado. Qualquer conteúdo incorretamente convertido ou elementos que a ferramenta tenha dificuldade de extrair automaticamente, podem ser corrigidos, sendo possível inseri-los manualmente na janela de edição. Antes de baixar os dados XML, o conteúdo criado é verificado automaticamente de acordo com as diretrizes de formato de dados XML estabelecidas pelo J-STAGE. Depois de realizar esse trabalho de edição, os dados XML completos são baixados e servem ao arquivamento de metadados de periódicos indexados pelo J-STAGE.

8. CONCLUSÃO

A disseminação do conhecimento científico através de periódicos passou por significativas mudanças ao longo do tempo. Desde a troca das cartas entre sociedades científicas até a formalização do periódico científico moderno enquanto canal formal de comunicação, diversos fatores mostraram como os processos comunicacionais estão em constante mudança e como os periódicos científicos não se mantiveram alheios a essas transformações. Desde a sua transição ao meio eletrônico eles vêm absorvendo novas tecnologias computacionais em seus processos editoriais.

A ampliação do acesso à informação científica na internet não foi possível sem que houvesse inovações e desenvolvimento de sistemas de informação. Dentre os avanços tecnológicos direcionados a esse propósito, o surgimento das linguagens de marcação teve papel fundamental. Estas por sua vez, também se transformaram ao longo do tempo e estão cada vez mais orientadas ao valor semântico e estrutural das informações contidas nas publicações eletrônicas.

O âmbito da produção e publicação de periódicos envolve diversos agentes: autores e pesquisadores que expressam novos conhecimentos através de artigos; editores envolvidos na revisão; bibliotecas e bases de dados que armazenam e disponibilizam, e: profissionais da tecnologia e da informação que buscam soluções para melhorar processos de tratamento da informação (estes muitas vezes trabalhando em conjunto). Todos eles, de uma forma ou de outra, são impactados pelas transformações tecnológicas.

Embora as linguagens de marcação baseadas em SGML estejam presentes nos processos editoriais de periódicos científicos desde a década de 1980, a exemplo da AAP Electronic Manuscript Standard - AAP DTD, ISO 12083 - Serial DTD, Modular Application for journals - MAJOUR e Simplified SGML for Serial Headers - SSSH DTD que foram abordadas neste estudo o que se observa é que a busca por aperfeiçoamentos e métodos que permitam melhorias nos processos editoriais parece ser uma constante entre desenvolvedores, editores e publicadores de periódicos envolvidos na comunicação científica.

O SGML foi considerado um bom padrão para publicação e arquivamento, mas a sua complexidade e altos custos, foram fatores que inibiram sua ampla adoção por

editores e publicadores. O HTML é um formato simples e uma linguagem bastante presente nas publicações eletrônicas, podendo ser associado a outras linguagens e folhas de estilos voltadas a apresentação das informações, contudo suas características pensadas para promover melhorias de apresentação de páginas web se mostraram muito limitadas para o gerenciamento de documentos e grandes volumes informacionais. O XML surgiu como uma tecnologia capaz de sanar as limitações de seus dois antecessores. Além de melhorias na apresentação dos conteúdos na página WEB, o XML foi desenvolvido centrado na estruturação de dados e maior flexibilidade, e pode ser considerado uma boa opção tanto para estruturar como compartilhar informações na web.

Do ponto de vista técnico, padrões extensíveis, em especial o XML, permitem além da descrição de recursos informacionais e criação de formatos derivados, maior interoperabilidade entre plataformas distintas, capacidade de arquivamento e reuso de dados. Há ampla literatura sobre padrões consolidados e linguagens de marcação, especialmente no caso do XML, que reforçam sua aplicação nestes processos. Porém, são escassas as que abordam sua aplicação especificamente a artigos científicos e as razões pelas quais os periódicos passaram a agregá-las nos processos de produção e publicação.

A publicação eletrônica possui um grande potencial informacional e as linguagens de marcação, em especial o XML, tiveram grandes contribuições. A inserção deste recurso computacional vem nas últimas décadas ganhando cada vez mais espaço na publicação científica, permitindo que sejam exploradas possibilidades de usos da informação pouco aplicáveis em outros formatos mais estáticos.

Bases de dados, como bibliotecas e portais acadêmicos, consideram importante receber dados de artigos de várias fontes em um único formato, pois isso favorece inúmeros processos de tratamento da informação.

Questões como, interoperabilidade, manejo e verificação automática de dados parecem ser os principais fatores que fizeram com o XML fosse mais fortemente agregado aos processos editoriais de periódicos eletrônicos. A recuperação da informação através da estruturação e agregação de valor semântico é pontuada como característica fundamental na literatura geral sobre linguagens de marcação, embora

este aspecto tenha sido pouco ressaltado nas motivações que levaram ao uso do XML nos processos de publicação de periódicos.

A JATS permite um grau elevado de flexibilidade por fornecer uma base padrão para editores de periódicos que podem adotar esse modelo pronto para uso ou com modificações personalizadas para resolver uma ampla gama de problemas de intercâmbio, acelerar o tempo de publicação e produção e reduzir custos de projetos do que o fariam se tivessem que desenvolver uma DTD específica para cada necessidade.

A JATS, possui características estruturais e descritivas dos elementos e atributos orientados especificamente a periódicos, tornando possível o processamento exaustivo de metadados por programas de computador, podendo assim servir a essas demandas de forma otimizada.

Por ser elaborada em um formato extensível, editores que optam por variações do padrão JATS, criam adequações para atender a necessidades específicas em seus processos de publicação. Além disso, a adoção da JATS mostra-se como uma possibilidade de aumentar a visibilidade dos periódicos através da indexação em bases internacionais que têm este padrão como requisito técnico para indexação.

As linguagens de marcação, enquanto padrões abertos e extensíveis, ao propiciar novas possibilidades de acesso e uso da informação eletronicamente podem ser consideradas importantes ferramentas orientadas às práticas bibliotecárias. A necessidade de descrição, armazenamento e intercâmbio de informações online, representam fatores essenciais que motivaram sua utilização em bases de dados e bibliotecas eletrônicas e, embora presentes desde 1980 e aprimoradas desde então à estas finalidades, observa-se na literatura nacional poucos trabalhos (PACKER, 1998, 2014) que abordam esses aspectos, sugerindo assim um campo de estudo ainda a ser mais explorado pela Biblioteconomia no Brasil.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMS, S.L.; ROSENBLUM, B. XML for e-journal archiving. **OCLC Systems & Services**, v. 19, n.4, p.155–161, 2003 Disponível em: <https://www-emerald.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/10650750310508126/full/pdf?title=xml-for-ejournal-archiving>. Acesso em 22 dez. 2020.
- ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação, Brasília** v. 31, n. 2, p. 5-13, 2002. Disponível em: <https://brapci.inf.br/index.php/res/v/18589>. Acesso em: 09 abr. 2020.
- BARNES, J.H. One giant leap, one small step: Continuing the migration to electronic journals. **Library Trands**, v. 45, n. 3, p. 404-415, 1997. Disponível em: http://apps-webofknowledge.ez67.periodicos.capes.gov.br/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=6Bq79BVryBtAGZuHjIS&page=1&doc=1. Acesso em: 06 jan. 2021.
- BAX, M. P. Introdução às linguagens de marcas. **Ciência da Informação, Brasília**, v. 30, n. 1, p. 32-38, 2001. Disponível em: https://www.scielo.br/scielo.php?pid=S0100-19652001000100005&script=sci_abstract&lng=pt. Acesso em: 10 abr. 2020
- BECK, J. NISO Z39.96 The Journal Article Tag Suite (JATS): What happened to the NLM DTDs? **The journal of electronic publishing: JEP** v. 14, n.1, p.106, 2011. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/22140303/>. Acesso em: 14 ago. 2020.
- BECK, J.; RANDALL, L. NLM DTD to NISO JATS Z39.96-2012. In: **The NCBI Handbook [Internet]** (2nd ed.). National Center for Biotechnology Information (US) (2013). Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK169004/>. Acesso em 22 dez. 2020.
- BECK, J.; USDIEN, T. B. JATS - A new standard for an old specification. **Information Standards Quarterly**, v. 25, n. 19-21, 2013. Disponível em: http://www.niso.org/apps/group_public/download.php/10418/SP_Beck_Usdin_JATS_isqv25no1.pdf. Acesso em: 21 dez. 2020.
- BROWN, A. XML in serial publishing: Past, present and future. **OCLC Systems and Services**, v.19, n.4, p.149–154, 2003 Disponível em: <https://www-emerald.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/10650750310698775/full/pdf?title=xml-in-serial-publishing-past-present-and-future>. Acesso em: 22 dez. 2020
- CORREA, P.C. Avaliação da usabilidade do portal de periódicos eletrônicos científicos da UNICAMP. 2019. 251 f. Dissertação (Mestrado em Divulgação Científica e Cultural) - Universidade Estadual de Campinas, Campinas. Disponível em: http://repositorio.unicamp.br/bitstream/REPOSIP/334091/1/Pereira_PaulaCorrea%20M.pdf. Acesso em: 08 dez 2020.
- CAVE, F. Article metadata standards: an historical review. **OCLC Systems and Services**, v. 19, n, 4, 2003, p. 144–148. Disponível em: <https://www-emerald.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/10650750310698766/full/pdf?title=article-metadata-standards-an-historical-review>. Acesso em: 20 dez. 2020.
- DEITEL, H. M. **XML: como programar**. Porto Alegre: Bookman, 2003.
- ENTRALGO-AMARO, R.; AGUIAR- CEDEÑO, J. D. Ventajas de la tecnología xml dentro de las ciencias de la información. **Revista Ciencias de la Información (Cuba)**,

v. 40, n. 3, p. 65-74, 2009. Disponível em: <<https://brapci.inf.br/index.php/res/v/59835>>. Acesso em: 17 abr. 2020.

FURGERI, S. O papel das linguagens de marcação para a ciência da informação. **Transinformação**, v. 18, n. 3, p. 225-239, 2006. Disponível em: <https://brapci.inf.br/index.php/res/v/115868>. Acesso em: 04 dez. 2020

GUÉDON, J-C. Sustaining the 'Great Conversation': the future of scholarly and scientific journals. In: PHILLIPS, A., COPE, B. (eds) **The Future of the Academic Journal**. 2 ed. Chandos Publishing, 2014. p. 85-136.

HUH, D. Journal Article Tag Suite 1.0: National Information Standards Organization standard of journal extensible markup language. **Science editing** v.1, n.2, p. 99-104, 2014. Disponível em: <https://www.esccienceediting.org/upload/se-1-2-99.pdf>. Acesso em: 31 out. 2020

INERA. E-journal archive dtd feasibility study, 2001. Disponível em: <https://jats.nlm.nih.gov/docs/Harvard-Archive-DTD-Feasibility-Study.pdf>. Acesso em: 12 dez. 2020

KANE, J.; MAKUCK, J.R. An overview of Electronic Publishing and Extensible Markup Language (XML). **Journal of Applied Communications**, v.82, n.2, p.1-15, 1988. Disponível em: <https://newprairiepress.org/cgi/viewcontent.cgi?article=2135&context=jac>. Acesso em: 23 dez. 2020

KATOH, K.; KOBAYASHI, T.; Kitazawa M. Reducing costs and expanding XML submissions with PDF to JATS conversion. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2012, **Anais eletrônicos...** Bethesda, National Center for Biotechnology Information, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK100490/>. Acesso em: 25 dez. 2020.

LAPEYRE, D. Introduction to JATS (Journal Article Tag Suite) [online], 2018 Disponível em: <https://www.xml.com/articles/2018/10/12/introduction-jats/>. Acesso em: 22 dez. 2020.

LANCASTER, F.W. The evolution of electronic publishing. **Library Trends**, v. 43, n. 4, p. 518-527, 1995. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F7924273E6948CFCF6C79011BFEFFD83?doi=10.1.1.204.8059&rep=rep1&type=pdf>. Acesso em: 19/10/2020

MARTINS, P.G.M. **Evolução das tecnologias de representação: das linguagens de marcação aos dados interligados**. 2018. 102. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de São Carlos, São Carlos, 2018. Disponível em: https://repositorio.ufscar.br/handle/ufscar/10491?locale-attribute=pt_BR. Acesso em: 22 dez. 2020.

MEADOWS, A. J. **A comunicação científica**. Brasília, Brinquet de Lemos, 1999.

MIRANDA, D.B.; PEREIRA, M. N. F. A revista científica como veículo de comunicação: uma revisão de literatura. **Ciência da Informação**, v. 25, n. 3, p. 375-382, 1996. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/462/421>>. Acesso em: 14 maio 2020.

MORAIS, E. F.; SOARES, M. B. **Web semântica para máquinas de busca**. In: SEMINÁRIO DA DISCIPLINA PROJETO E ANÁLISE DE ALGORITMOS, 2003, Pampulha. **Anais eletrônicos...** Pampulha: PPGCC/UFMG. Disponível em:

<https://homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>. Acesso em: 06 jan. 2021

MURRAY-RUST, P.; RZEPA, H. S. Scientific publications in XML - towards a global knowledge base. **Data Science Journals**, v.1, n1. p. 84–98, 2003. Disponível em: https://www.jstage.jst.go.jp/article/dsj/1/0/1_0_84/pdf/-char/en. Acesso em: 02 dez 2020

OLIVEIRA, E. B. P. M. periódicos científicos eletrônicos: definições e histórico. *Informação e Sociedade*, v.18, n.2 p. 69-77, 2008.

PACKER, A.L, SciELO: uma metodologia para publicação eletrônica. **Ciência da informação, Brasília**, v. 27, n. 2, p. nd, 1998. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200002&lng=en&nrm=iso. Acesso em: 24 dez. 2020.

PACKER, A., et al. XML, por quê? [online]. *SciELO em Perspectiva*, 2014. Disponível em: https://blog.scielo.org/blog/2014/04/04/xml-porque/#.X_MB6RZv--o. Acesso em: 22 dez. 2020.

PARKIN M. How JATS supports data integration: extracting data availability statements and funding information from research articles in Europe PMC. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2019, **Anais eletrônicos...** Bethesda, National Center for Biotechnology Information, 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK541158/#parkin-SD1>. Acesso em: 29 nov. 2020

RAMALHO, R. A. S.; MARTINS, P. G. M.; SOUSA, J. L. Evolução das linguagens de marcação: um breve histórico à luz da área de ciência da informação. **Informação@Profissões**, v. 6, n. 2, p. 20-34, 2017. Disponível em: <https://brapci.inf.br/index.php/res/v/61470>. Acesso 05 abr. 2020.

ROSENBLUM, B.D. Smoke screens of impossibility. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2015, **Anais eletrônicos...** Bethesda, National Center for Biotechnology Information, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK285911/>. Acesso em: 20 dez 2020.

ROSENBLUM, B.; GOLFMAN, I. A decade of DTDs and SGML in scholarly publishing: What have we learned? (2002) In: *Extreme Markup Languages 2002*. **Anais eletrônicos...** Montréal, Computer Science, 2002. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.139.2374&rep=rep1&type=pdf>. Acesso em: 06 dez 2020.

SANTA ANNA, J. Mapeamento sistemático na base de dados dm Ciência da Informação: Periódicos científicos em discussão”. **Brazilian Journal of Information Science: Research Trends**, v. 12, n. 1, p. 68-79, 2018. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/7296/5008>. Acesso em: 05 dez. 2020.

SANTOS, S. **SciELO 15 Anos - Preparação de textos em XML, HTML, PDF e Epub**, 2015. Disponível em: https://www.youtube.com/watch?v=813OtRY_l9s. Acesso em: 18 dez. 2020.

SAYÃO, L. F. Repositórios digitais confiáveis para a preservação de periódicos eletrônicos científicos. **Ponto de Acesso, Salvador**, v. 4, n. 3, p. 68-94, 2010.

Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/4709/0>. Acesso em: 12 dez 2020.

SOUZA, M. F, S.; Periódicos Científicos Eletrônicos: apresentação de modelo para análise de estrutura. 2002 142 f. Dissertação (Mestrado em Ciência da informação) - Universidade Estadual Paulista, UNESP. Disponível em: <https://repositorio.unesp.br/handle/11449/93718>. Acesso 10 jan. 2021.

SCIELO. Documentação para a DTD SciELO - Versão 3.1. Disponível em: <http://old.scielo.org/php/level.php?lang=pt&component=56&item=20>. Acesso em: 24 dez. 2020.

SCIELO. Guia de uso de elementos e atributos XML para documentos que seguem a implementação SciELO Publishing Schema, 2019. Disponível em: https://scielo.readthedocs.io/projects/scielo-publishing-schema/pt_BR/master/. Acesso em: 24 dez. 2020.

STUMPF, R. C. Passado e futuro das revistas científicas. **Ciência da Informação**, v.25, n.3, 1996. Disponível em: <https://brapci.inf.br/index.php/res/v/19165>. Acesso em: 15 abr. 2020.

TARGINO, M. G. Comunicação científica na sociedade tecnológica: periódicos eletrônicos em discussão. **Comunicação e Sociedade**, n. 31, p.71-98, 1999. Disponível em: <https://www.metodista.br/revistas/revistas-ims/index.php/CSO/article/view/7890/6620>. Acesso em: 12 nov. 2020

TARGINO, M. G. Comunicação científica: uma revisão de seus elementos básicos. **Revista Informação & Sociedade: Estudos**, Joao Pessoa, v. 10, n. 2, p. 37-85, 2000. Disponível em: <https://brapci.inf.br/index.php/res/v/92195>. Acesso em: 06 jan. 2020.

WUSTEMAN, J. XML and e-journals. **OCLC Systems and Services**, v.19, n4, p.125–131, 2003a. Disponível em: <https://www-emerald.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/10650750310508081/full/pdf?title=xml-and-ejournals>. Acesso em: 18 dez. 2020.

WUSTEMAN, J. XML and e-journals: The state of play. **Library Hi Tech**, v.21, n1, p21–33, 2003b. Disponível em <https://www-emerald.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/07378830310467373/full/pdf?title=xml-and-ejournals-the-state-of-play>. Acesso em: 18 dez. 2020.