

---

# Abordagem sobre vocabulários controlados para arquivos: conceitos, aplicações e metodologias

*Observations on controlled vocabularies for archives: concepts, applications and methodologies*

---

**Walter Moreira (1), Luciana Davanzo (2), Isabela Santana de Moraes (3)**

(1) Universidade Estadual Paulista (UNESP), Av. Hygino Muzzi Filho, nº 737, CEP: 17525-000, Marília, SP, Brasil, walter.moreira@marilia.unesp.br

(2) luciana.davanzo@gmail.com

(3) isabela@marilia.unesp.br

## Resumo

A utilização de vocabulários controlados em unidades arquivísticas é tomada como objeto de estudo dessa pesquisa, cujo objetivo é identificar e descrever essa temática no interior da literatura periódica científica da arquivologia. Foram utilizadas as seguintes bases de dados para o estabelecimento do corpus documental: SCOPUS, BRAPCI e LIST&A. O período de cobertura considerado foi de dez anos e a estratégia de busca utilizada contemplou os termos “tesauro”, “vocabulário controlado” e “arquivologia”. A leitura analítica do corpus foi realizada com subsídios da técnica de análise de conteúdo. Os resultados apontam a relativa incipiência do tema na arquivologia. Observou-se também que ainda é necessário uma aproximação maior da arquivologia com os vocabulários controlados, uma vez que, majoritariamente a literatura remete-se à biblioteconomia. A ausência de padronização na descrição de assunto e a baixa oferta de soluções pela literatura científica especializada tem se convertido nos principais obstáculos para as ações de interoperabilidade requeridas aos sistemas de informação modernos.

**Palavras-chave:** Arquivologia. Organização da informação arquivística. Representação da informação arquivística. Vocabulários controlados.

## 1. Introdução

A arquivologia, enquanto campo de investigação e de práticas relacionadas à produção, organização, guarda, preservação e utilização das informações arquivísticas, tem sido cada vez mais desafiada a oferecer soluções eficazes e sintonizadas com os processos contemporâneos de tratamento e disseminação da informação, marcados pela presença das tecnologias de informação e comunicação.

Nesse cenário e tendo em vista o documento arquivístico digital (Levy, 1994; Rondinelli, 2011), as questões teóricas e metodológicas

## Abstract

The use of controlled vocabularies in archives units is taken as an object of study of this research, which aims to identify and describe this subject within the scientific periodical literature of archives science. The following databases were used to establish the documentary corpus: SCOPUS, BRAPCI, and LIST&A. The coverage period considered was ten years and the search strategy included the terms "thesaurus", "controlled vocabulary" and "archives science". The analytical reading of the corpus was carried out with grants from the content analysis technique. The results show the relative incipient theme in archives science. It was also noted that it is still necessary to promote further approximation of archives science with controlled vocabularies, since, mostly literature refers to the library science. The lack of standardization in the subject description and the low supply of solutions for specialized scientific literature has been converted into the main obstacles to the interoperability of cases brought to modern information systems.

**Keywords:** Archives science, Information organization, Information retrieval, Controlled vocabularies.

impostas como problemas científicos para a consolidação epistemológica da Arquivologia, ganham novos contornos.

O conceito de documento arquivístico, como é próprio dos conceitos, *lato sensu*, ainda não é absolutamente pacífico. Há uma longa tradição nessa discussão que envolve pesquisadores como Casanova (1928), Schelenberg (1956), Duranti, Eastwood e Macneil (2002) e Rondinelli (2011), entre outros.

Rondinelli (2011), por exemplo, aponta a divergência que existe entre os termos “documento arquivístico” e “informação arquivística”, sendo o termo “documento arquivístico”, mais comum

em países de língua inglesa e o termo “informação arquivística” mais comumente utilizado em países de língua latina.

Independentemente da denominação, entretanto, para que sejam considerados como documentos de arquivo, esses documentos precisam ser produzidos e recebidos por uma determinada instituição. O documento arquivístico (ou a informação arquivística), portanto, já possui em sua gênese um objetivo. Num primeiro momento serve à própria instituição que o produziu e, posteriormente, pode assumir caráter histórico e, desse modo, servir aos pesquisadores ou aos usuários da sociedade civil, de modo geral.

Os documentos arquivísticos são produzidos, reunidos e organizados a partir de um contexto específico de produção, seja o produtor desse contexto, uma pessoa, uma família ou uma instituição (pública ou privada). Os documentos estabelecem um conjunto de relações entre si que precisam, portanto, ser mantidas. Para que sejam localizados e recuperados é preciso que sejam definidos os pontos de acesso correspondentes, isso é, é preciso descrevê-los em função dos interesses do sistema onde se localiza e também de outros sistemas com os quais espera-se que dialogue.

O conceito de informação e, naturalmente, o conceito de informação arquivística, tradicionalmente assentados em concepções epistemológicas desenhadas para modelos analógicos também precisam ser repensados em função do cenário hipertextual contemporâneo (Lamarca Lapuente, 2006). Em comum tem-se que ambos os cenários são marcados pela presença e pela intermediação da linguagem.

Nesse ambiente hipertextual, em que se confundem produtores e usuários da informação e no qual a própria configuração conceitual do documento é objeto de constantes revisões, indicadores de credibilidade, precisão, exatidão e racionalidade da informação são ainda mais requeridos.

Nesse cenário, as formas de produzir, acessar e recuperar a informação mudaram de maneira significativa. Não é mais absolutamente necessário, por exemplo, que o pesquisador ou usuário esteja no mesmo local físico em que os documentos estão custodiados, isso pode ocorrer por meio do ambiente *web*, por exemplo (Buckland, 1999). Os documentos, inclusive, podem não estar fisicamente armazenados no local de acesso, pois nesse cenário a duplicação dos documentos, para fins de acesso, é um processo com custo extremamente baixo. Há, evidentemente, outros motivos, alheios a essa pesqui-

sa, pelos quais se pode questionar e fazer objeção à duplicação.

Naturalmente, propostas modernas de organização da informação devem considerar os recursos tecnológicos disponíveis. A base tecnológica influencia a produção, a circulação e aumenta a diversidade de suportes documentais. Essa mesma base, contudo, favorece o desenho de instrumentos mais precisos de representação e de recuperação da informação.

Há, como se pode depreender, questões de fundo que subjazem às questões tecnológicas propriamente ditas. O problema básico que configura os campos da arquivologia, da biblioteconomia, da documentação e da ciência da informação, nesse prisma, ainda vigora: para que uma informação possa cumprir seu papel no ciclo de produção, organização, disseminação e geração de novo conhecimento, isso é, para que seja socializada, é necessário que seja localizável e recuperável pelo usuário, o que é possível por meio de tratamento linguístico informacional adequado.

A produção de informação arquivística interpõe-se como intermediária entre a produção e o uso do conhecimento. Essa intermediação ocorre, basicamente, por meio da linguagem do sistema, o que requer consistência terminológica como condição necessária ao diálogo eficaz entre a representação da informação pela unidade arquivística (entrada) e representação da informação pelo usuário (saída). As designações “entrada” e “saída” nesse caso consideram obviamente o ponto de vista do sistema. Considerando-se o ponto de vista do usuário essas designações serão, naturalmente, invertidas.

Não é possível falar em consistência terminológica em ações de representação nas quais são envolvidos vários sujeitos, com diferentes formações discursivas, sem algum recurso da padronização que lhes garanta o acordo. Deve-se atentar, contudo, para o caráter mediador, não coercitivo, da padronização que se pode alcançar com o uso de vocabulários controlados consensualmente compartilhados.

Observando-se esse caráter mediador dos vocabulários controlados, ele é requerido, de modo geral, em todas as ações que envolvem tratamento, organização e recuperação da informação. Desse modo, também desempenham um papel vital em instrumentos mais *modernos* como ontologias, as folksonomias e os recursos de *linked data*.

Em suma, se há um processo em que se requer a descrição de elementos não contemplados pelas normas arquivísticas e mais suscetíveis

às variações culturais, institucionais e terminológicas, como é o caso da descrição do assunto do documento, é necessário considerar a função de centralidade que o vocabulário controlado assume (Buckland, 1999; Lancaster, 2002).

Desse modo, a utilização de vocabulários controlados em unidades arquivísticas é tomada como objeto de estudo, considerando-se que esses instrumentos de organização do conhecimento oferecem subsídios não somente para a organização como também para a recuperação da informação, independentemente de suporte analógico ou digital.

Um vocabulário controlado, como resultado do processo de controle vocabular utilizado num determinado sistema de informação, é uma linguagem construída para fins documentários, constituída de termos organizados em estrutura relacional. A utilização de um vocabulário controlado visa a padronizar e facilitar a entrada e a saída de dados em um sistema de informação documentária, promovendo eficácia na comunicação entre os usuários e o sistema. A estrutura relacional revela as relações hierárquicas, associativas e de equivalência que os conceitos guardam entre si.

Um vocabulário controlado permite representar a informação e o conhecimento por meio de um conjunto padronizado e finito de termos para facilitar a recuperação (Kobashi, 2008; Buckland, 1999).

Observando-se a relevância teórico-prática do tema e tendo-se em vista sua relativamente baixa abordagem na literatura mais específica da arquivologia, objetiva-se com essa pesquisa identificar e analisar a abordagem (conceitos, aplicações e metodologias) sobre vocabulários controlados e sua aplicação em unidades arquivísticas, conforme sua expressão na literatura científica da arquivologia e da ciência da informação.

Dentre os objetivos da arquivologia, é sempre importante lembrar, está a criação de condições para proporcionar acessibilidade aos documentos. Desenvolver e utilizar instrumentos que contribuam para esse fim torna-se, portanto, fundamental. Nessa conjuntura, acredita-se que os vocabulários controlados podem contribuir de maneira bastante satisfatória, pois se não houver acordo terminológico não será possível a precisão na comunicação.

Os vocabulários controlados são elaborados de maneira a facilitar e tornar mais preciso o acesso à informação, portanto, sua elaboração leva em consideração alguns fatores, tais como, os autores, o catalogador, a estrutura sindética, o

pesquisador e a consulta formulada" (Buckland, 1999).

Os vocabulários controlados são elaborados a partir de necessidades específicas de cada instituição, assim, cada vocabulário controlado será único de acordo com cada instituição e em função de suas particularidades. Entretanto, independentemente das especificidades de cada instituição, o vocabulário controlado será elaborado a partir de algumas vertentes fundamentais, conforme explicita Buckland (1999): os vocabulários do autor, do catalogador, da estrutura relacional, do pesquisador e da expressão de busca.

Para que um vocabulário controlado se torne efetivamente útil é preciso que seja organizado de modo a refletir tanto os objetivos do sistema de informação para o qual foi elaborado, isso é a linguagem do sistema, quanto a linguagem dos usuários (Kobashi, 2008). Desta forma, sua construção requer trabalho integrado, envolvendo tanto os gerenciadores do sistema quanto os usuários da informação. O vocabulário controlado deve servir ao sistema, o que implica dizer que sua aplicação ocorre tanto no processo de representação para fins de organização da informação, realizados pelo indexador ou arquivista, quanto em processos de representação para fins de recuperação da informação, realizados pelo usuário.

Os termos, num vocabulário controlado, são dispostos em ordem hierárquica e alfabética (macroestrutura e microestrutura). A macroestrutura é o eixo temático global, em que se apresentam as categorias mais amplas dos campos de conhecimentos abrangidos de forma hierárquica. A microestrutura dispõe, em forma alfabética os termos do vocabulário (Kobashi, 2008).

Considerando-se que o vocabulário controlado é formado a partir das linguagens de especialidade, isso é das terminologias específicas dos campos que pretende organizar e representar, e considerando-se que essas terminologias formam subsistemas dentro do conjunto geral da língua, é natural que os vocabulários controlados sejam instrumentos dinâmicos, que requerem revisões constantes.

O vocabulário controlado é utilizado para indicar o significado de uma determinada expressão, mas, sendo a língua é instável, as palavras podem assumir novos significados, independentemente daqueles que um lexicógrafo possa ter "fixado" (Buckland, 1999). No discurso, ou seja, na utilização efetiva da língua é que surgem diferenças, ambiguidades e incertezas de significado.

O vocabulário controlado também é importante para questões de identidade, que, por sua vez, são indispensáveis para as instituições e para a sociedade, pois o uso da linguagem é inevitavelmente baseado em cultura (Buckland, 1999).

### 1.1. Procedimentos metodológicos

A pesquisa tem caráter bibliográfico, exploratório e descritivo. A escolha desse delineamento para a pesquisa justifica-se em função de se ter verificado, em estudo anterior (Davanzo; Moreira, 2014), a insuficiência do tratamento do tema relativo aos vocabulários controlados na literatura brasileira de ciência da informação.

Desse modo, para o estabelecimento do *corpus documental 1*, foram utilizadas as seguintes bases de dados: SCOPUS (Elsevier); Base Referencial de Artigos de Periódicos de Ciência da Informação (BRAPCI) e *Library, Information Science & Technology Abstracts with Full Text (EBSCO)*. A escolha das bases considerou sua representatividade, confiabilidade e cobertura, atestadas em trabalhos como Jacso (2005) e Bufrem, Costa, Gabriel Júnior, Pinto (2010) e Grácio e Oliveira (2012).

O período de cobertura considerado foi de dez anos (2005-2014) e a estratégia de busca utilizada contemplou os termos “tesauro”, “vocabulário controlado” e “arquivologia”, considerando-se também suas variações de equivalência sinonímica, de número e de idioma (inglês, português e espanhol).

Observou-se, como critério de seleção para a composição do *corpus documental 2*, a disponibilidade do artigo em texto completo e, preenchida essa condição, a presença (simultânea ou não) dos termos de busca nos campos de título, palavras-chave e resumos.

Foram identificados inicialmente 23 artigos na SCOPUS, 21 artigos na LISTA e apenas um artigo na BRAPCI. Após a eliminação dos itens duplicados e dos artigos que traziam as expressões apenas nos campos de resumo, obteve-se um total geral de 28 artigos.

Para fins de identificação dos artigos relevantes para a composição do *corpus de análise*, os 28 artigos foram pontuados e organizados num *ranking*. Desse modo, conforme a ocorrência da palavra nos diversos campos de busca, o artigo obteve as pontuações indicadas no Quadro I.

Campo(s) de ocorrência	Valor
Palavra-chave	3
Título	2
Resumo	1
Palavra-chave + título + resumo	6
Palavra-chave + Título	5
Palavra-chave + resumo	4
Título + resumo	3

Quadro I. Critérios de pontuação dos artigos para composição do *corpus de análise*

Considerando-se que a pontuação mínima obtida foi 3 e a máxima foi 12, estabeleceu-se como nota de corte a média de 6 pontos. Desse modo, o *corpus de análise* foi composto de 14 artigos.

A leitura analítica dos textos completos selecionados foi realizada com subsídios da técnica de análise de conteúdo, conforme a proposta de Bardin (2009).

## 2. Resultados e discussão

Conforme apontado nos objetivos, interessaram à análise desenvolvida nessa pesquisa os artigos que apresentam a conjunção de tratamento dos temas “vocabulários controlados” e “arquivos”. Interessou saber como essas temáticas tem se aproximado.

A ordenação dos trabalhos conforme o critério de pontuação apresentado na subseção 1.1 (procedimentos metodológicos) está apresentada no Quadro II.

Referência	Pontuação
Schenkolewski-Kroll e Tractinsky (2006)	12
Neal (2008)	10
Nilbe & Tarkpea (2014)	10
Siqueira (2011)	10
Eito Brun (2011)	9
Balkan e Bell (2014)	7
Baum e Kempf (2012)	7
Jenkins, Oppenheim, Probeta & Hubbard (2008)	7
Lin, Murray, Dorr, Hajic & Pecina (2009)	7
Reid & Simmons (2007)	7
Young & Mandelstam (2013)	7
Clarke (2008)	6
Gimenez Chornet (2011)	6
Jenkins, Probeta, Oppenheim & Hubbard (2007)	6

Quadro II. Artigos ordenados conforme a pontuação recebida

É importante observar a quantidade relativamente baixa de trabalhos recuperados, considerando-se que foram consultadas três bases de dados de periódicos, sendo duas internacionais, com cobertura de dez anos.

Apenas um trabalho (Schenkolewski-Kroll & Tractinsky, 2006) apresentou a ocorrência dos termos (“tesauro” ou “vocabulário controlado” e “arquivo”) em todos os campos (título, resumo e palavra-chave) e obteve pontuação máxima conforme os critérios estabelecidos.

Dentre os quatorze artigos que compuseram o *corpus* de análise, havia três que, após leitura analítica e categorização, verificou-se que abordavam temática não exatamente relacionada aos objetivos desta pesquisa. Desse modo, foram incluídos na categoria “não se aplica”: Neal (2008) e Young e Mandelstam (2013), que abordam a aplicação de tesouros em bibliotecas, e Nilbe e Tarkpea (2014), que trata da aplicação de tesauro em um repositório.

Seguindo-se a técnica de análise de conteúdo (Bardin, 2009), foram identificados os temas e estes foram organizados em categorias semânticas definidas *a posteriori*, ou seja, após a análise do material.

Recomenda-se, na aplicação da análise de conteúdo (Bardin, 2009), que as categorias sejam autoexcludentes. Essa recomendação não foi observada com rigor por dois motivos: a) o pequeno número de trabalhos identificados e b) a identificação de mais de um tema (categoria) nos trabalhos. A definição do foco temático do artigo ocorreu, portanto, em conformidade com os interesses pragmáticos específicos dessa investigação. Além do mais, a categorização, ensina a mesma Bardin (2009), não deve ser encarada de modo descontextualizado e se reveste sempre de uma hipótese de organização.

Sendo assim, foram analisados 11 artigos, distribuídos nas seguintes categorias de análise: a) conceitos: trabalhos mais conceituais, com foco de abordagem mais teórica nas relações entre vocabulários controlados e arquivos; b) aplicações: trabalhos que priorizam a descrição de aplicações de soluções, isso é, que apresentam casos de sucesso da aplicação de vocabulários controlados em arquivos; c) metodologias: trabalhos que oferecem mais detalhes a respeito das metodologias utilizadas para o desenvolvimento e aplicação de vocabulários controlados em arquivos.

Os trabalhos foram agrupados conforme apresentado no Quadro III e a discussão é apresentada nas subseções subsequentes.

<b>Categoria</b>	<b>Autor(es)</b>
Trabalhos conceituais	Clarke (2008), Lin, Murray, Dorr, Hajic e Pecina (2009), Siqueira (2011), Baum e Kempf (2012), Balkan e Bell (2014)
Foco na aplicação	Reid e Simmons (2007), Eito Brun (2011), Baum e Kampf (2012), Balkan e Bell (2014)
Foco nas metodologias	Schenkolewski-Kroll e Tractinsky (2006), Jenkins, Oppenheim, Proberts & Hubbard (2007), Jenkins, Oppenheim, Proberts & Hubbard (2008), Giménez-Chornet e Escrig-Giménez (2011)
Não se aplica	Neal (2008), Young (2013), Nilbe (2014)

Quadro III. *Categorização do corpus de análise*

### 2.1. Trabalhos conceituais

Nessa categoria foram incluídos os seguintes trabalhos: Clarke (2008), Lin, Murray, Dorr, Hajic e Pecina (2009), Siqueira (2011), Baum e Kempf (2012), Balkan e Bell (2014).

Os vocabulários controlados facilitam o acesso a coleções digitais e outras pela representação explícita dos princípios subjacentes de organização (Lin, Murray, Dorr, Hajic & Pecina, 2009)

Um conceito associado ao conceito de vocabulário controlado que parece próximo (se não o estiver fato) do consensual é o conceito de interoperabilidade. Isso está previsto, inclusive, em normas como a recentemente revisada ISO 25964, que orienta a elaboração de tesouros. Dividida em duas partes, a primeira refere-se às características gerais dos tesouros e dos elementos que o compõem, a segunda parte trata especificamente da interoperabilidade entre vocabulários controlados (International Organization for Standardization [ISO], 2011).

A ISO 25964 privilegia o conceito em detrimento das palavras ou termos. A mesma norma, em versão anterior, considerava o tesauro como um conjunto de termos interconectados por uma rede de relações semânticas. A proposta atual define o tesauro como um conjunto de conceitos relacionados semanticamente. Ainda que a tipologia de relações seja a mesma de um tesauro baseado nas relações entre termos, a terminologia sintetiza uma diferença substancial de concepção, mais alinhada aos fundamentos da teoria comunicativa da terminologia.

Num tesauro baseado na relação entre termos, as relações são definidas entre os termos. Nesse tipo de relação podem-se relacionar termos que denotam um mesmo conceito, como ocorre, por exemplo, com as abreviaturas. Num tesauro

baseado em conceito, algumas relações ocorrem entre os conceitos e outras entre os termos. (Pastor-Sanchez, Martínez Mendez & Rodríguez-Muñoz, 2009; Clarke & Zeng, 2012).

Entre os autores que destacam a necessidade de aspectos de interoperabilidade nos vocabulários controlados estão Balkan e Bell (2014).

A preocupação com os aspectos de interoperabilidade em vocabulários controlados decorre dos crescentes esforços de internacionalização em ambientes científicos. Baum e Kempf (2012) consideram, contudo, que a normalização necessária à indexação internacional coerente, a qual é vital para a recuperação da informação, ainda não está suficiente desenvolvida. Os autores não fazem referência a nenhuma norma de modo específico, não se sabe se tinham conhecimento da norma ISO citada anteriormente ou se ela está incluída na crítica.

De qualquer modo, os processos de internacionalização, pelo que carregam de complexidade cultural, andam a passos lentos. Esforços no sentido de promover a padronização e internacionalização existem desde, pelo menos, as propostas da documentação, realizadas por Paul Otlet. Baum e Kempf (2012) situam os trabalhos sobre essas propostas nos anos 1970 e atribuem a falta de diretrizes coerentes à ausência de pesquisas sobre os diferenças nos modelos de indexação de resultados de pesquisa realizados pelos diversos países.

Clarke (2008) ratifica o importante avanço verificado por meio do conjunto de metadados *Dublin Core* e chama a atenção para o papel que os vocabulários controlados podem assumir no preenchimento do elemento “assunto”.

O único trabalho brasileiro localizado, conforme a estratégia de busca estabelecida para essa investigação e apresentada nos procedimentos metodológicos, é de Siqueira (2011). Este trabalho, analisado apressadamente, poderia ser colocado na categoria “aplicações”. Sua leitura atenta, contudo, revela que, apesar de analisar um arquivo específico, o Sistema de Arquivos da USP (SAUSP), isso é realizado com o objetivo de demonstrar a eficácia do vocabulário controlado como instrumento que agrega precisão na descrição e na localização da informação arquivística.

## 2.2. O foco nas aplicações

Nessa categoria foram incluídos os seguintes trabalhos: Reid e Simmons (2007), Eito Brun (2011), Baum e Kampf (2012) e Balkan e Bell (2014).

O tesouro multilingue *European Language Social Science Thesaurus* (ELSST) foi objeto de dois artigos: Baum e Kampf (2012) e Balkan e Bell (2014). O ELSST é administrado pelo *UK Data Archive*, que também administra o tesouro monolíngue *Humanities and Social Science Electronic Thesaurus* (HASSET), do qual o ELSST é derivado.

Eito Brun (2011) descreve a experiência de aplicação de padrões desenvolvidos em XML para a gestão e publicação das descrições e documentos do fundo Engenheiro Carlos Fernández Casado. O trabalho foi desenvolvido para o Centro de Estudos Históricos de Obras Públicas e Urbanismo do Ministério do Desenvolvimento, na Espanha.

A solução envolveu o desenvolvimento de um ambiente para edição e criação de instrumentos de descrição *Encoded Archival Description* (E-AD) compatíveis com as normas internacionais ISAD(G) E ISAAR(CPF). Este ambiente permite acesso a vocabulários controlados locais ou remotos para a realização da descrição.

Reid e Simmons (2007) apresentam a experiência de aplicação de arquivos de autoridade e tesouros na definição dos pontos de acesso do *Archival Research Catalog* (ARC). O ARC é um projeto desenvolvido pela *National Archives and Records Administration* (NARA), no âmbito do governo federal dos Estados Unidos. O uso de vocabulário controlado, relatam os autores, assegura ao ARC consistência e precisão na recuperação da informação.

## 2.3. Foco nas metodologias

Nessa categoria foram incluídos os seguintes trabalhos: Schenkolewski-Kroll e Tractinsky (2006), Jenkins, Oppenheim, Probets & Hubbard (2007), Jenkins, Probets, Oppenheim & Hubbard (2008), Giménez-Chornet e Escrig-Giménez (2011).

Giménez-Chornet e Escrig-Giménez (2011) descrevem uma aparentemente complexa experiência de integrar quatro unidades arquivísticas espanholas. Para isso, treze especialistas em arquivo utilizaram o mesmo servidor para criar descrições padronizadas com recurso das normas ISAD(G) E ISAD(CPF) e de três tesouros para fins de indexação, definindo instituições, lugares geográficos e assuntos.

Este trabalho descreve os aspectos de construção de um protótipo de um tesouro incorporado num sistema de busca situado na intranet que atende a quatro arquivos históricos geograficamente separados.

Jenkins, Oppenheim, Proberts e Hubbard (2008), descrevem o processo de criação de um vocabulário controlado para analisar processos de autoarquivamento de acordos de transferência de direitos autorais realizados por editores de revistas. O vocabulário controlado consiste de três categorias que possibilitam descrever: o que pode ser autoarquivado, as condições e as restrições do autoarquivamento. O vocabulário controlado, na avaliação dos autores do trabalho, mostrou-se eficaz, mas requer revisão e atualização contínuas em função da dinamicidade inerente ao assunto que cobre.

Dentre as dificuldades encontradas durante o desenvolvimento do vocabulário controlado, relatam a necessidade de reduzir o número de descritores em função da grande variedade de termos utilizados pelos editores, sem consistência e sem precisão.

O primeiro estágio na criação do vocabulário, que não seguiu, segundo informam os autores, um método formal, foi familiarizar-se com as políticas de autoarquivamento e com os acordos de transferência de direitos autorais. A partir dessa análise, os termos foram colhidos e organizados nas seguintes categorias: “que” (*pre-prints, pos-prints etc.*), “onde” (repositório público, página pessoal do autor etc.), “qual versão” (versão do autor, versão formatada do editor etc.) e “requisitos” (o que o autor deve fazer para que o autoarquivamento seja permitido pelo editor).

Além do artigo descrito anteriormente, no qual Jenkins, Oppenheim, Proberts & Hubbard (2008) falam sobre a criação de vocabulário controlado para a análise de acordos de transferência de direitos autorais, denominado *RoMEO studies 7*, os mesmos autores, em estudo publicado um ano antes, discutem aspectos lógicos do autoarquivamento, denominado *RoMEO studies 8* (Jenkins, Oppenheim, Proberts & Hubbard (2007).

O artigo de Schenkolewski-Kroll e Tractinsky (2006) foi o que melhor respondeu à estratégia de busca utilizada para localizar material para essa pesquisa, conforme já apontado anteriormente. É também o que apresenta mais claramente a relação entre a descrição de material arquivístico, indexação e recuperação da informação, bem como aspectos sobre a construção de tesouros para fins arquivísticos.

Esse trabalho relata a experiência desenvolvida nos arquivos israelenses, bem como a situação geral dos arquivos nesse país em função das normas internacionais e sua adaptação às necessidades arquivísticas nacionais.

Os autores descrevem dois projetos específicos que representam diferentes visões: o tesouro do *National Insurance Institute of Israel* e os arquivos *Religious Zionism at Barllan University*. Em conclusão, o artigo destaca uma proposta de criação de um sistema de descrição e recuperação da informação arquivística baseado em ISAD(G), ISAAR(CPF) e um tesouro geral adequado a todos os arquivos israelenses.

### 3. Considerações finais

Os resultados apontam para a relativa incipiência do tema na arquivologia e sinalizam uma possível confusão conceitual em relação ao termo “vocabulário controlado”, tomado, em alguns casos, como sinônimo de “tesouro”.

Observou-se, ainda, a ausência de um referencial teórico forte e específico sobre a aplicação de vocabulários controlados em arquivos, já que, normalmente, os textos são construídos com o referencial teórico e o modelo de abordagem da biblioteconomia em relação a esses instrumentos. A esse respeito é preciso desenvolver novas e mais profundas investigações.

Embora pareça predominar na literatura arquivística a denominação “vocabulários controlados”, a literatura recente da *International Society of Knowledge Organization* (ISKO) tem adotado de formal preferencial o termo “knowledge organization system” (KOS), como mais generalizante.

Essa investigação é limitada pelo recorte que estabelece, como, de fato ocorre, aliás, com qualquer outra investigação. Investigações mais amplas devem incluir termos como “sistemas de organização do conhecimento”, “instrumentos de organização do conhecimento”, “linguagens documentárias”, “ontologias”, “taxonomias”, “folksonomias” e “classificação facetada” e suas relações com a arquivística. É preciso verificar também a confusão terminológica que se estabelecem entre essas expressões.

A pesquisa que motivou esse relatório foi desenvolvida com a função precípua de organizar e analisar o material bibliográfico (artigos de periódicos) disponível sobre as relações entre vocabulários controlados e arquivos. Desse modo, pretende-se que sirva como guia para a condução de futuras investigações sobre o tema.

As normas ISAD(G) e ISAAR(CPF) orientam a descrição de material arquivístico com base na proveniência e auxiliam na descrição dos fundos em vários níveis. Falta, ainda, uma norma ou um conjunto de diretrizes que sejam aplicadas à descrição e indexação de conteúdo dos regis-

tros. A ausência de padronização nesse sentido, bem como a baixa oferta de soluções pela literatura científica especializada tem se tornado um dos principais obstáculos para as ações de interoperabilidade requeridas aos sistemas de informação modernos. Dificultam, outrossim, como atestado pelos diversos estudos citados, a precisão na representação e na recuperação da informação arquivística.

## Referências

- Balkan, Lorna; Bell, Lucy. (2014). Linking thesauri - ELSST as hub for social science data terms. *IASSIST Quarterly*, 16-21.
- Bardin, L. (2009). Análise de conteúdo. Lisboa: Edições 70.
- Baum, K., Kempf, A. O. (2012). Thesaurus-based indexing of research data in the social sciences: opportunities and difficulties of internationalization efforts. *IASSIST Quarterly*, 12-15.
- Buckland, M. (1999, May). Vocabulary as a central concept in library and information science. In T. Arpanac et al. (Eds.). *Digital libraries: interdisciplinary concepts, challenges, and opportunities: proceedings of the Third International Conference on Conceptions of Library and Information Science*. Dubrovnik, Croatia. 3-12.
- Bufrem, L. S., Costa, F. D. O.; Gabriel Júnior, R. F. & Pinto, J. S. P. (2010). Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. *Perspectivas em Ciência da Informação*, 15(2), 22-41.
- Casanova, E. (1928). *Arquivística*. Arti Grafiche Lazzeri.
- Clarke, S. G. D. (2008). The last 50 years of knowledge organization: a journey through my personal archives. *Journal of Information Science*, 34(4), 427-437.
- Davanzo, L. & Moreira, W. (2014). O vocabulário controlado como ferramenta do processo de organização e recuperação da informação. *Anais do XIV Encontro Nacional de Pesquisa em Ciência da Informação*. Belo Horizonte: ECI/UFMG. 1142-1148.
- Clarke, S. G. D. & Zeng, M. L. (2012). From ISO 2788 to ISO 25964: the evolution of the thesaurus standards towards interoperability and data modeling. *Information Standards Quarterly*, 24(1), 20-26.
- Duranti, L; Eastwood, T; Macneil, H. (2002). *Preservation of the integrity of electronic records*. Dordrecht: Kluwer Academic.
- Eito Brun, R. (2011). Aplicabilidad de EAD en archivos históricos de ingeniería: el fondo Carlos Fernández Casado. *Scire*, 17(1), 35-39.
- Giménez-Chornet, V. & Escrig-Giménez, M. (2011). Designing a thesaurus to give visibility to the historical archives in the Archivo del Reino in Valencia. *Knowledge Organization*, 38(2), 154-166.
- Grácio, M. C. C. & Oliveira, E. F. T. (2012). Visibilidade dos pesquisadores no periódico *Scientometrics* a partir da perspectiva brasileira: um estudo de cocitação. *Em Questão*, 18(edição especial), 99-113.
- International Organization for Standardization (2011). *ISO 25964: information and documentation- Thesauri and interoperability with other vocabularies*. Genebra: ISO.
- Jacso, P. (2005). As we may search: comparison of major features of the Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537-1547.
- Jenkins, C., Oppenheim, C., Proberts, S. & Hubbard, B. (2008). RoMEO studies 7: creation of a controlled vocabulary to analyse copyright transfer agreements. *Journal of Information Science*, 34(3), 209-307.
- Jenkins, C., Proberts, S., Oppenheim, C. & Hubbard, B. (2007). RoMEO studies 8: self archiving: the logic behind the colour-coding used in the copyright knowledge bank. *Program*, 41(2)124-133.
- Kobashi, N. Y. (2008). Vocabulário controlado: estrutura e utilização. [Texto preparado para subsidiar o debate sobre vocabulário controlado da pesquisa Mapeamento para a reunião da Rede de Escolas de Governo em 15 e 16 dez. 2008].
- Lamarca Lapuente, M. J. (2006). *Hipertexto: el nuevo concepto de documento en la cultura de la imagen*. Tesis de doctorado en Fundamentos, Metodología y Aplicaciones de las Tecnologías Documentales y Procesamiento de la Información, Universidad Complutense de Madrid. Recuperado em 30 jan. 2015, de <http://www.hipertexto.info>.
- Lancaster, F. W. (2002). *El control del vocabulário en la recuperación de información*. Valencia: Universitat de Valencia.
- Levy, D. M. (1994). Fixed or fluid? Document stability and new media. In: *Proceedings of the European Conference on Hypermedia Technology (ECHT 94)*. Edinburgh, UK. *Proceedings...* Edinburgh, p. 24-31.
- Lin, J., Murray, G. C., Dorr, B. J., Hajic, J. & Pecina, P. (2009). A cost-effective lexical acquisition process for large scale thesaurus translation. *Lang Resources & Evaluation*, n. 43, p. 27-40, 2009.
- Neal, D. (2008). News photographers, librarians, tags, and controlled vocabularies: balancing the forces. *Journal of Library Metadata*, 8(3), 199-219.
- Nilbe, S. & Tarkpea, T. (2014). Using the estonian subject thesaurus in the digital environment. *Cataloging & Classification Quarterly*, 52(1), 32-41.
- Pastor-Sanchez, J. A., Martínez Mendez, F. J., Rodríguez-Muñoz, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14(4).
- Reid, L. J. E. & Simmons, C. J. (2007). Authority control at the National Archives and Records Administration. *Journal of Archival Organization*, 5(1/2), 95-120.
- Rondinelli, R. C. (2011). *O conceito de documento arquivístico frente à realidade digital: uma revisitação necessária*. Tese de doutorado em Ciência da Informação, Universidade Federal Fluminense, Brasil, RJ, Niterói.
- Schenkolewski-Kroll, S. & Tractinsky, A. (2006). Archival description, information retrieval, and the construction of thesauri in israeli archives. *Archival Science*, 6, 69-107.
- Schellenberg, T. R. (1956). *Modern archives: principles and techniques*. Chicago: University of Chicago Press.
- Siqueira, J. C. (2011). Recursos linguísticos para análise de vocabulário controlado: o caso do SAUSP. *Biblionline*, 7(2), 52-62.
- Young, J. L.; Mandelstam, Y. (2013). It takes a village: developing Library of Congress genre/form terms. *Staff Publications*, Paper 2.