

# **Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: análise da Plataforma Lattes**

**Fábio Mascarenhas Silva**

**Professor Adjunto do Departamento de Ciência da  
Informação da UFPE**

**Johanna Wilhelmina Smit**

**Docente junto ao Departamento de  
Biblioteconomia e Documentação da**

*Discussão, avaliação e apresentação de parâmetros para a organização da informação científica e tecnológica brasileira, enfocando os problemas do acesso à informação em sistemas abertos, especificamente a Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Desenvolveu-se um estudo exploratório para identificar se a natureza aberta do sistema compromete a consistência dos dados na recuperação da informação, concluindo-se que há comprometimento. A partir da sistematização dos resultados, apresentam-se sugestões para aprimorar o sistema.*

**Palavras-chave:** *Sistemas abertos; Sistemas de Recuperação de Informação; Informação científica e tecnológica - Brasil; Plataforma Lattes; Organização da informação científica e tecnológica.*

## **Information organization in open electronic systems of Scientific and Technological Information: analysis of the Lattes Database**

*Discussion, assessment and presentation of parameters for organization of the Brazilian Scientific and Technological Information are shown with focus on the problems of access to information in open systems,*

*specifically the Lattes Database of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). An exploratory study is presented which was developed with CVs taken from the Lattes Database. The goal was to identify whether the open nature of the system could put at risk the consistency of the data when information is retrieved. Since the answer was affirmative we present suggestions to improve the system from a perspective of results systemization.*

**Keywords:** *Information organization; Open systems; Information retrieval systems; Brazilian scientific and technological information; Lattes Data Base.*

*Recebido em 16.08.2008 Aceito em 12.03.2009*

## 1 Introdução

O objeto de análise desta pesquisa foi a Plataforma Lattes (PL) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) em que, mais especificamente, explorou-se o sistema de gestão de currículos denominado Currículos Lattes. A PL representa a experiência do CNPq na integração de bases de dados de currículos e de instituições da área de Ciência e Tecnologia. Sua base de dados já ultrapassou 1 milhão de registros (currículos) e as informações constantes da PL servem tanto às atividades de gestão, como no apoio à formulação de políticas para a área de Ciência e Tecnologia. A pesquisa sobre a PL foi conduzida a partir de um referencial teórico relacionado à organização da informação em meio eletrônico. As leituras críticas dos fundamentos teóricos nesta temática partiram do objetivo maior desta pesquisa: investigar se há comprometimento da consistência dos dados nos sistemas abertos de informação. A motivação para investigar tal assunto surgiu da percepção da crescente tendência de os próprios pesquisadores alimentarem os dados nos Sistemas de Informação Científica e Tecnológica (SICT) brasileiros.

O objetivo geral da pesquisa foi, portanto, discutir, avaliar e propor sugestões à organização da Informação Científica e Tecnológica (ICT) brasileira em meio eletrônico, caracterizada pela livre inserção de dados. Para alcançar esse objetivo foi necessário cumprir algumas etapas, quais sejam: traçar um retrospecto histórico da ICT brasileira; analisar criticamente os recursos voltados à organização da informação; desenvolver um estudo exploratório em um SICT nacional, a PL, com o propósito de identificar se há comprometimento na consistência dos dados decorrentes da natureza aberta do sistema; relacionar os procedimentos de organização da informação utilizados pela PL com recursos tradicionalmente utilizados para o tratamento da informação, como os vocabulários controlados, a fim de propor melhorias.

Partiu-se da hipótese de que a atual metodologia adotada para coleta e organização da informação na PL, ainda que elaborada a partir de estruturas computacionais bem definidas, pautadas em ontologias e linguagens de marcação, seja insuficiente para proporcionar uma organização da informação consistente e confiável. Tal problema compromete o processo de Recuperação da Informação, e também a geração e o uso dos dados da PL para apoio na gestão da C&T.

Um dos pilares desta pesquisa é uma tendência que vem ganhando força: a facilidade de os próprios autores produzirem não apenas o conhecimento propriamente dito, mas também a representação desse conhecimento nos SICT. Essa mudança, visível em recursos como os *arquivos abertos*, periódicos científicos, ou em grandes sistemas de ICT como a PL, se por um lado facilita a disponibilização/acesso aos documentos, por outro pode prejudicar o processo de Recuperação da Informação e o uso dos dados para a produção de indicadores em C&T.

Como o foco deste trabalho é a ICT, ressalta-se a preocupação de autores diante do fato de que a aceitação dos recursos eletrônicos para a produção/disponibilização da ICT é um processo ainda a ser assimilado pela comunidade científica. Outra preocupação diz respeito à facilidade e flexibilidade de os usuários inserirem, além dos documentos, as representações de suas produções científicas e técnicas, ou seja, seus metadados. Essas representações tanto dizem respeito às descrições físicas como temáticas dos documentos digitais. Há discussões na literatura brasileira da Ciência da Informação sobre o assunto. É o que se vê em recente publicação de Marcondes (2006), ou em um outro trabalho - com participação do mesmo autor - no qual são debatidas as novas formas de cooperação em ICT (MARCONDES; SAYÃO, 2002). Salienta-se que, nesse último, a referida cooperação condiz com recursos de interoperabilidade entre sistemas de informação que dependem de coincidências sintáticas entre conteúdos; ou seja, são pré-definidas relações de equivalência entre campos e seus respectivos atributos.

Apesar dos grandes avanços alcançados pela PL nos últimos anos, ainda é preciso aperfeiçoar as mediações deste sistema com os usuários, e imagina-se que uma alternativa seja adotar mecanismos de controle adequados aos princípios de organização da informação. A ausência de tais mecanismos sugere que não se previu, na etapa de planejamento da PL, que o preenchimento dos currículos seria feito por uma comunidade bastante heterogênea e nem sempre familiarizada com recursos de informação.

## **2 Análise da Plataforma Lattes**

Os principais objetivos da PL são: fornecer subsídios para elaborar políticas públicas ou diagnósticos da C&T brasileira e informar o currículo de pesquisadores, ou seja, o que os mesmos fazem e fizeram. Para que esses objetivos sejam efetivamente alcançados, é necessário introduzir mecanismos de controle na etapa de inserção dos dados, no

processamento desses dados e na forma de apresentação dos currículos. Para sugerir aprimoramentos no sistema, analisou-se a PL em duas etapas: a primeira, a partir da lógica dos arquivos pessoais, e a segunda, observando os procedimentos de preenchimento do sistema. Em ambas foram feitas análises críticas, porém, na segunda, foram também intercaladas sugestões direcionadas à PL que podem ser incorporadas a qualquer outro SICT. Por fim, são apresentadas discussões finais e sugestões que completam estas análises.

### **3 A Plataforma Lattes e a lógica dos Arquivos Pessoais**

Uma das funcionalidades da PL é gerar currículos que serão tornados públicos. Esses currículos são documentos que, à moda da PL, organizam referências a documentos (alguns públicos e outros privados) do arquivo pessoal, ou institucional, dos cientistas. Desta forma, na PL o currículo é um documento que deveria refletir a relação entre os documentos/atividades dos usuários cadastrados.

Estes currículos servem a um delimitado segmento de atuação social: o segmento dos atores da C&T brasileira. Cada parte da estrutura dos currículos descreve atuações ou produções em C&T e cada usuário cadastrado preenche, individualmente, as atividades que foram por ele desenvolvidas. No caso das atividades desenvolvidas com a participação de outros autores, é possível citá-los, entretanto será necessário que cada participante citado descreva, em seus respectivos currículos, a atividade comum a todos eles. Na percepção arquivística, um arquivo pessoal idealmente organizado requer uma análise das atividades realizadas pela pessoa da qual se organizará os documentos, os quais serão organizados e agregados em função das atividades exercidas pela pessoa que os acumulou ao longo da vida.

Mas a lógica da PL não prioriza este mesmo princípio, pois distribui as atividades em função do que as mesmas representam, cada uma delas entendida isoladamente, desconsiderando o contexto no qual foram realizadas. Assim, uma palestra cadastrada na PL é somente uma palestra, o que impossibilita contextualizá-la no ambiente de uma pesquisa em curso ou em uma atividade de extensão que tenha gerado o convite para a palestra. Um exemplo prático: um pesquisador desenvolveu, ao longo de dois anos, uma metodologia para o uso de indicadores bibliométricos na formulação de políticas públicas em C&T. Seu trabalho resultou na publicação de dois artigos e um livro, em convites para proferir palestras em um congresso e um simpósio, e ainda um convite para ser consultor num programa de capacitação para técnicos do Ministério da Ciência e Tecnologia.

As atividades do exemplo citado permitem o registro de ao menos sete atividades do pesquisador: um processo, dois artigos publicados, um livro, um curso de curta duração (com material didático) e duas palestras. Todas são desdobramentos decorrentes da criação de uma metodologia específica. Ao preencher o currículo, as atividades são desmembradas de

um núcleo de ação que originou um conjunto de ações, ou seja, são descontextualizadas. Vê-se que a menção a cada uma dessas atividades remove das mesmas seu significado, pois elas não foram geradas de forma descontextualizada, mas sim dentro de um contexto bem definido. E ainda, o registro de cada atividade, isoladamente, contribui para tornar o currículo demasiadamente longo e pouco informativo, em decorrência da falta de contextualização das atividades arroladas.

O princípio de categorização da PL é baseado numa visão dualista, que entende as atividades como produção ou atuação. Essa divisão define as regras de relações entre as classes e hierarquias que compõem a PL. Numa visão arquivística, tais relações desfavorecem a constituição de arquivos pessoais, pois segrega em partes o que em vida se realizou de forma articulada ou contextualizada.

Essa divergência da PL com relação aos princípios arquivísticos compromete um dos objetivos da plataforma, que é o de apresentar em formato organizado e padronizado os currículos dos pesquisadores. Como conseqüência, o currículo de um pesquisador experiente e com uma produção representativa, sob o ponto de vista quantitativo, pode ser bastante extenso. Quanto mais informações no currículo, mais difícil será fazer uma leitura sistêmica dele. Por sistêmica referimo-nos a uma análise conjuntural da vida do pesquisador. Desta forma, a PL, ao ignorar a lógica arquivística, prejudica a compreensão da atuação do pesquisador e, neste sentido, compromete o objetivo mencionado no início deste parágrafo.

A base da PL, apesar de por definição ser constituída por documentos biográficos de atores da C&T, não se caracteriza como uma fonte ideal para a organização de documentos que retratam as atividades de pesquisadores. A forma como foi desenvolvida privilegia uma distribuição das atividades exercidas pelo sujeito de forma isolada e descontextualizada, tornando necessário registrar repetidas vezes um conjunto de ações que, originalmente, ocorreram de forma concatenada. Isso resolve um problema: o registro de todas as ações. Mas acarreta outros: a produção de um documento biográfico demasiadamente longo e pouco informativo.

## **4 Análise do preenchimento da Plataforma Lattes**

A forma de cadastro e o respectivo preenchimento dos campos da PL foram modificados ao longo dos anos e continuam em processo de aperfeiçoamento. Dito isso, é importante entender que as interfaces de preenchimento disponíveis no período das análises<sup>1</sup> podem sofrer alterações com o passar do tempo. De todo modo, buscou-se convergir as avaliações críticas aos aspectos estritamente relacionados à organização da informação.

Em junho de 2007 a PL mantinha sete módulos, são eles: *Dados Gerais, Produção Bibliográfica, Produção Técnica, Orientações, Produção Cultural, Eventos, e Bancas*. Cada módulo contém diferentes campos que

---

<sup>1</sup> Análises feitas no período de junho a agosto de 2007.

permitem ao usuário inserir conteúdos em forma de texto. Para fins desta pesquisa optou-se por categorizar as formas de preenchimento da PL em três grupos: Autonomia Total, Autonomia Parcial, e Sem Autonomia. Apesar de o sistema não ser assim subdividido, tal classificação é proposta para sistematizar as análises e discussões. As características essenciais de cada categoria são:

- Autonomia Total: O usuário tem a liberdade de cadastrar as palavras que desejar, sem restrição ou qualquer direcionamento. Ex: Título de uma publicação;
- Autonomia Parcial: São campos em que, inicialmente, se tem autonomia total, porém cada novo termo cadastrado pelo usuário é automaticamente armazenado no sistema, que vai criando uma lista de termos exclusiva do usuário. Nas vezes seguintes em que o usuário inserir outros termos, será possível consultar e adotar termos anteriormente criados por ele. O usuário pode excluir qualquer termo dessa lista, ou incluir novos;
- Sem Autonomia: O sistema prevê, inicialmente, opções que o usuário deve selecionar. Entretanto, a existência dessas opções prévias não impede que novos termos sejam incluídos, caso o usuário não se satisfaça com as opções oferecidas. Ex.: Áreas do Conhecimento, Setores de aplicação.

Para a análise de cada categoria utilizaram-se currículos consultados na PL que, apesar de não terem sido coletados segundo procedimentos de amostragem, não foram aleatoriamente escolhidos, mas sustentados pela estratégia desenvolvida pela SciELO. Para aspectos da organização da informação, postulou-se que as partes do currículo referentes à produção bibliográfica fossem mais apropriadas, pois exigem representações conceituais mais complexas se comparadas, por exemplo, com dados pessoais do pesquisador.

Pensando na possibilidade de usar outro sistema como parâmetro, para fins comparativos, optou-se por utilizar exemplos retirados de periódicos disponíveis na SciELO. A escolha deu-se em razão da credibilidade alcançada por este sistema na comunidade científica brasileira (e internacional). Para a seleção dos periódicos, considerou-se a quantidade de fascículos já publicados, fator esse que evidencia a consolidação do periódico perante os pares. A escolha foi feita a partir da lista denominada COLEÇÃO DA BIBLIOTECA, que apresenta os periódicos disponibilizados na SciELO, ordenados em oito categorias. Para cada categoria foi selecionado um periódico e, com isto, as categorias e respectivos periódicos assim foram arrolados:

1. Ciências Agrárias (Arquivo Brasileiro de Medicina Veterinária e Zootecnia);
2. Ciências Biológicas (Memórias do Instituto Oswaldo Cruz);
3. Ciências da Saúde (Arquivos Brasileiros de Cardiologia);
4. Ciências Exatas e da Terra (*Brazilian Journal of Physics*);

5. Ciências Sociais Aplicadas (Ciência da Informação);
6. Engenharias (*Brazilian Journal of Chemical Engineering*);
7. Lingüística, Letras e Artes (DELTA);
8. Humanas (Estudos Avançados).

Para uniformizar o período dos artigos usados na análise, definiu-se primeiro o número dos periódicos publicados no ano de 2006; ou seja, ao final foram analisados 8 números de periódicos, sendo o primeiro número de 2006 de cada um dos 8 selecionados. Ao todo, a análise contemplou 80 artigos, publicados por 282 autores.

#### a) Análise dos Campos com Autonomia Total:

Para a análise da categoria dos campos com Autonomia Total, consideraram-se inconsistentes os dados preenchidos nos currículos de forma diferente da produção bibliográfica registrada no periódico da SciELO. Para verificar este aspecto, considerou-se na PL o campo título como o mais adequado, pois, sob o ponto de vista sintático, só pode haver equivalência entre dois registros de títulos quando ambos forem idênticos, diferentemente de uma avaliação semântica, na qual um mesmo significado pode estar presente em termos diferentes.

Observou-se em alguns casos que o título original do artigo estava em português, mas que os autores o cadastraram em inglês na PL. Para fins de recuperação da informação, as palavras cadastradas em inglês apenas serão úteis para estratégias de buscas formuladas com termos na língua inglesa. É importante ressaltar que as buscas feitas na PL normalmente o são em língua portuguesa, ou seja, se o título original do artigo estiver em inglês (algo comum na literatura estrangeira e em alguns casos da brasileira também), haverá comprometimento nos resultados.

Foi comum verificar que, tanto no periódico como no currículo, há autores que optam por registrar seus nomes próprios de formas diferentes. Percebeu-se, também, que foi recorrente encontrar artigos na SciELO de autores que eram cadastrados na PL, não tendo sido criado o enlace na página do artigo do periódico. Um dos motivos desta falha se explica certamente pela diferença nos nomes dos autores. Confrontando-se dados de periódicos da SciELO com os currículos dos pesquisadores autores dos respectivos artigos, identificou-se, na categoria dos campos com Autonomia Total, no preenchimento: erros de digitação, o uso do idioma inglês (quando o sistema majoritariamente adota a língua portuguesa), e até mesmo a ausência ou troca nos títulos.

#### b) Análise dos Campos com Autonomia Parcial:

Trata-se de campos inicialmente sem opções (similar aos campos com Autonomia Total), em que cada novo termo cadastrado é armazenado no sistema. Para a análise dos campos com Autonomia

Parcial foram considerados aspectos de sinonímia e homonímia, que são representações lingüísticas diferentes para objetos iguais ou similares, o que demonstra a natureza semântica destes campos. Para a análise foram confrontadas as palavras-chave cadastradas pelos autores na PL com as palavras-chave registradas nos artigos publicados nos periódicos disponíveis na SciELO.

É patente a adoção de termos no plural. Sabe-se que, para fins documentários, a normalização gramatical é preconizada para sistemas de informação, prevendo evitar divergências na grafia das palavras, muito embora diferenças na grafia ocasionadas pelo uso do plural/singular não interfiram em alguns sistemas de buscas que identificam a ausência da letra "S" no final da palavra.

Uma situação identificada na categoria dos campos com Autonomia Parcial, que ocorre também na dos campos com Autonomia Total, é o uso de termos em língua diferente do português. Um dos motivos que conduz os autores a cadastrarem as palavras em outro idioma é que os artigos, mesmo publicados no Brasil, estão em outra língua. É previsível que os autores usem a língua adotada na publicação, porém, notou-se um fato curioso: os autores utilizam a mesma língua, mas não necessariamente repetem as mesmas palavras-chave usadas no artigo.

Utilizar mais de um idioma para criar palavras-chave não seria algo novo para os pesquisadores, pois as normas para publicações científicas já exigem resumos e palavras-chave em pelo menos uma língua diferente (normalmente em inglês). Tal procedimento poderia ser adotado também para os campos de título e palavras-chave na PL.

Entre as inconsistências em um sistema de informação, a dispersão de termos é uma das mais comprometedoras. Os problemas acarretados envolvem aspectos de natureza tecnológica (como a agilidade do sistema devido à extensa lista de termos no banco de dados, o que influencia na rapidez da resposta do sistema) e também elementos relacionados a princípios de organização e tratamento da informação, redundando em problemas na Recuperação da Informação. Para este estudo entender-se-á que a "dispersão" resulta da diversidade de palavras-chave usadas para representar uma dada produção científica, ou seja, da ausência de controle na inserção de palavras-chave na PL, por parte dos autores de cada artigo. Essa diversidade acarreta a "pulverização" da informação.

Tradicionalmente, as palavras-chave usadas em documentos servem como representações temáticas dos próprios documentos. Tais representações não almejam completar a mensagem, mas sim oferecer um recurso auxiliar para a recuperação da informação. Na PL, as palavras-chave da produção científica tanto podem ser utilizadas no processo de recuperação dos currículos, como também proporcionar estudos métricos da produção científica brasileira e respectivos indicadores de C&T.

Tanto para fins de recuperação da informação como de estudos métricos, é importante que - além do planejamento do sistema para essas finalidades - a base de dados de currículos seja alimentada a partir de uma orientação voltada a esses propósitos. Mas os exemplos



demonstraram o contrário. A característica aberta da PL permite que o preenchimento dos campos seja realizado à mercê da percepção que os usuários alimentadores têm do seu funcionamento ou dos objetivos perseguidos no momento do preenchimento e dos objetivos da própria PL.

Era de se esperar inconsistências relativas à sinonímia e/ou homonímia nos campos de palavras-chave em sistemas abertos. No caso da PL, para o preenchimento de palavras-chave, não há esclarecimentos a respeito do uso dos termos; assim, é improvável que, desconhecendo princípios de organização da informação, os usuários se preocupem com questões de natureza documentária.

A atual quantidade de registros na PL torna possível aos seus administradores fazer um levantamento (a partir da base de dados do sistema) representativo dos termos mais utilizados no campo das palavras-chave. Se esse levantamento fosse realizado, é provável que houvesse viabilidade técnica para estratificar os termos segundo áreas de conhecimento dos currículos. A partir de uma relação dos termos mais adotados em cada área é viável implementar um recurso que auxilie o usuário a preencher os campos, sugerindo os termos mais adotados por seus pares. Ressalta-se que a finalidade não seria a de eliminar a especificidade, ou seja, aquilo que por ser menos freqüente pudesse ser mais informacional: o intuito é o de apresentar uma opção de grafia a partir das primeiras letras do termo que o usuário estivesse cadastrando.

No entanto, essa não é uma solução para as atuais inconsistências da PL no que diz respeito à organização da informação. Serviria somente como uma forma de orientação do sistema para o preenchimento dos campos, considerando-se que atualmente não há, com exceção da lista de termos criada pelo próprio usuário (que também pode conter sinonímias, formas gramaticais diferentes, etc.), indicação de quais palavras-chave o usuário poderia adotar. Isso permitiria ao usuário ter uma noção (quantitativa) dos termos mais adotados por seus pares. Na forma atual, o usuário, no momento do preenchimento, visualiza apenas os termos que ele próprio cadastrou.

É visível que as inconsistências na categoria dos campos com Autonomia Parcial da PL são prejudiciais à Recuperação da Informação. Mas a inconsistência que gera dispersão é desfavorável principalmente às análises conjunturais dos currículos, comprometendo a desejada formulação de indicadores de C&T a partir de dados da PL.

Notou-se, nos exemplos analisados da PL, que a representação usual dos pesquisadores segue uma tendência para o uso de uma linguagem natural. As relações entre palavras-chave de autores e periódico, e vice-versa, demonstraram que os procedimentos requeridos em sistemas fechados de informação são pouco usuais. Para uma análise aprofundada dos termos exemplificados seria necessário o domínio das áreas de conhecimento dos artigos publicados.

É importante entender que a linguagem controlada busca reduzir as variações semânticas e sintáticas de uma linguagem natural. A funcionalidade e o êxito da linguagem controlada – ou linguagem

documentária - limitam-se a ambientes de informação, com o objetivo de organizar e recuperar a informação; são linguagens construídas e, por isso, consideradas artificiais, não tendo aplicabilidade em outros ambientes e situações.

No que diz respeito ao controle de termos, a categoria dos campos com Autonomia Parcial da PL é um pouco menos crítica que a dos campos com Autonomia Total, pois possibilita o re-uso de termos já existentes no currículo do pesquisador. Novas e desnecessárias palavras-chave podem ser evitadas graças ao fato de o sistema gerar uma lista de termos que poderá ser consultada quando da atualização de um currículo, o que pode minimizar a inclusão de sinônimos, hipônimos e plurais. No entanto, esta possibilidade não basta para que o sistema proporcione uma recuperação da informação eficaz e tampouco seja capaz de produzir, a partir de sua base de dados, indicadores em C&T consistentes.

Interessantes reflexões de Kobashi e Santos (2007) dizem respeito a essas problemáticas da PL concernentes à produção de indicadores. Os autores explicam que os dados temáticos necessários para a produção de indicadores não podem ser tratados segundo as mesmas políticas de indexação para fins de Recuperação da Informação. A especificidade é o princípio básico aplicado na indexação para recuperação, que objetiva discriminar informação por meio da criação de classes constituídas por uma quantidade manejável de registros bibliográficos. Tal procedimento desfavorece estudos bibliométricos.

Em geral, os termos utilizados na indexação para recuperação proporcionam grande quantidade de classes de baixa frequência, resultando em um núcleo reduzido e em alta dispersão. Por outro lado, é preciso cautela na reformatação para que a substituição de termos específicos por níveis mais genéricos não gere classes com frequências muito altas, pois frequências altas tendem a não apresentar significados.

Há características da PL que indicam problemas de planejamento e operação relacionados às suas finalidades. No que tange ao planejamento, observou-se que a concepção, apesar de (supostamente) orientada ao desenvolvimento de um sistema para recuperação e geração de indicadores para Gestão de C&T, não previu - em campos importantes como os das palavras-chave - meios necessários a estes fins como, por exemplo, um vocabulário controlado ou uma árvore hierárquica de termos.

Segundo seus desenvolvedores (GRUPO STELLA, 2007), a PL seguiu um modelo em que os usuários "são produtores e multiplicadores de conhecimento (pesquisadores, docentes, estudantes, grupos de pesquisa, etc.)". Assim, o conjunto que configura a PL é baseado num princípio denominado "regras de negócio dos sistemas", onde cada um dos usuários utiliza e gera a informação que conformará o sistema. No entanto, ao optar por seguir a chamada regra de negócio dos sistemas, o desenvolvimento da PL priorizou a economia de custos, abrindo mão da sua qualidade.

O debate sobre “regras de negócio dos sistemas” está relacionado com os novos modelos de serviços da *Web*, abertos à participação dos usuários para o compartilhamento de serviços e informações. Na percepção de Catarino e Baptista (2007), trata-se de um novo paradigma para a organização dos conteúdos de recursos digitais na *Web* designados, genericamente, de *folksonomias*. Já na visão de Noruzi (2007), a folksonomia corresponde a uma taxonomia auto-gerada (no original *user-generated*) pelo usuário para que ele – o usuário – possa categorizar e recuperar conteúdos da *Web* a partir de etiquetas denominadas “*tags*”. De acordo com esse autor, as *tags* podem contribuir para a melhoria dos sistemas de busca da *internet*, em razão de os conteúdos categorizados formarem um vocabulário compartilhável entre usuários.

É precipitado considerar as folksonomias como um novo paradigma, pois a criação de etiquetas (*tags*) de marcação para conteúdos na *Internet*, por enquanto, não configura, sob o ponto de vista da organização da informação, um paradigma que possa ser expandido para todo e qualquer contexto informacional.

#### c) Análise dos Campos Sem Autonomia:

Nos campos sem autonomia o sistema oferece um conjunto de opções pré-cadastradas ao usuário. Em campos como “Áreas do Conhecimento”, “Setores de Atividade” e “Título do Periódico”, por exemplo, devem-se cadastrar preferencialmente itens pertinentes consultando os itens pré-cadastrados no sistema. Porém, é facultado incluir novos itens que não constem nesse conjunto de opções. Na página de busca avançada da PL há filtros que facilitam e refinam o processo de busca. Ao aplicar um ou mais filtros, o usuário aumenta as chances de harmonizar sua estratégia de busca com os registros da base, pois as opções oferecidas pelos filtros reproduzem as mesmas oferecidas aos usuários no preenchimento dos campos Sem Autonomia.

Para a análise dos campos Sem Autonomia foi observado o campo “Áreas do Conhecimento”. Contudo, diferentemente das categorias anteriores, não foram feitas comparações a artigos de periódicos da SciELO. Foram analisadas somente as representações de Áreas de Conhecimento que os autores fizeram a partir dos artigos disponíveis na SciELO. Dos três tipos de campos para preenchimento da PL, os campos Sem Autonomia constituem, sem dúvida, a opção mais restritiva no que diz respeito à liberdade de inserção do usuário. Ainda que exista a possibilidade de se incluir palavras diferentes das listadas, em nenhum currículo foi identificada alguma inclusão.

Nos campos da PL “Áreas do Conhecimento” e “Setores de Atividades”, as opções seguem uma estrutura hierárquica. As estruturas hierárquicas permitem uma visualização de níveis mais genéricos para os mais específicos. As Áreas de Conhecimento estão baseadas na Tabela de Áreas de Conhecimento do CNPq, enquanto que o campo Setores de

Atividades dispõe de uma classificação própria do sistema, que indica setores econômicos e sociais relacionados ao trabalho desenvolvido.

São perceptíveis as diferenças nas indicações das Grandes Áreas e, principalmente, Áreas de Conhecimento. Do ponto de vista do uso das informações da PL para a Gestão em C&T, isso pode significar que: há um conjunto fragmentado de dados pouco informativo para indicar comportamentos no âmbito da produção científica brasileira, ou então, quando pesquisadores de diferentes áreas produzem conjuntamente, acentua-se o caráter multidisciplinar/interdisciplinar de co-autorias. Em ambas as situações, a interpretação adequada dos dados exigirá uma rigorosa compreensão dessas nuances.

Quanto à Recuperação da Informação, o uso dos campos Sem Autonomia na PL é relativamente bem explorado para fins de busca de currículos. É oferecida a possibilidade de busca pela produção de acordo com as Áreas de Conhecimento. Como a estratégia de busca é formulada a partir de uma lista controlada, torna-se mais fácil estabelecer coincidências entre os termos definidos pelos usuários e os existentes na base do sistema.

Por fim, os motivos que conduzem os usuários a preencherem os campos Sem Autonomia com termos genéricos ou então com termos que representam coisas distintas (ex: FÍSICA – ARQUEOLOGIA) requerem investigações mais apropriadas e, para tanto, são necessários estudos de usuários focados na representação da informação. Não se deve desconsiderar que a raiz do problema pode também estar presente no recurso que é oferecido ao usuário. No caso específico do exemplo explorado, a árvore hierárquica talvez seja insuficiente para representar de forma exaustiva a diversidade de Áreas de Conhecimento.

Um fato curioso chama a atenção: a indicação da Área de Conhecimento por parte dos autores dos artigos do periódico Ciência da Informação foi a mais genérica de todas as áreas. É patente a escolha pela opção CIÊNCIAS SOCIAIS APLICADAS/CIÊNCIA DA INFORMAÇÃO. A limitação da amostragem analisada nesta pesquisa impede de se chegar a conclusões mais detalhadas sobre este fato, porém, arrisca-se dizer que os autores, diante da precariedade de representação da árvore de conhecimento, optaram por pecar por generalidade, evitando subdivisões mais específicas.

## **5 Discussões e Sugestões**

O planejamento/desenvolvimento da PL, intencionalmente ou não, desconsiderou as vantagens proporcionadas aos sistemas pelo controle que a eles podem ser atribuídos. Em contrapartida, foi beneficiado por um grande ganho econômico ao compartilhar com a comunidade acadêmica o compromisso de alimentar um sistema que serve de apoio aos órgãos de fomento brasileiros. Reduz-se o custo de investimento, mas, em compensação, perde-se consistência nas informações disponibilizadas. Princípio semelhante (apenas no aspecto de alimentação dos sistemas)

ocorre com os diversos repositórios abertos – também chamados de arquivos abertos - voltados ao ambiente da C&T. E, seguindo o mesmo princípio, o de tornar o usuário um agente ativo nas representações de conteúdos – existem as folksonomias que, entretanto, não se restringem ao universo da ICT.

Na PL, a perda de consistência na Recuperação da Informação, conforme visto nos exemplos analisados, poderia ter sido menor se fossem adotados procedimentos orientados ao controle do sistema. Tais procedimentos podem ser utilizados não somente nos sistemas de currículos, mas em todos os que utilizam termos para representação de informações. Desta forma, são apresentadas a seguir recomendações voltadas à organização da informação, que, apesar de já bastante difundidas nos domínios da Ciência da Informação e de não serem inéditas, podem contribuir para a concepção e o funcionamento de sistemas eletrônicos de informação.

O controle de vocabulário se inicia com procedimentos que Smit e Kobashi (2003) denominaram “micro” e que servem ao controle dos termos ou expressões em arquivos, tais como: a) Normalização gramatical, b) Opções de grafia, c) Controle de sinonímia e d) Controle de homonímia. O procedimento “macro” diz respeito à organização dos termos em formatos previstos em classificações ou tesouros. Cada procedimento acima enumerado será detalhado a seguir, iniciando-se com os procedimentos “micro”.

a) Normalização gramatical (recomenda-se a adoção da forma substantiva, masculina e singular dos termos):

É possível incluir em sistemas como a PL recursos similares aos utilizados em corretores ortográficos dos editores de texto. Isso evitaria, no mínimo, erros elementares de digitação. Mas, além de corrigir erros, o recurso seria mais proveitoso se funcionasse a partir de um vocabulário de termos criado para áreas específicas. Tal atitude seria imprescindível para possibilitar a identificação de termos adequadamente, levando-se em conta a adoção da forma no substantivo, masculina e singular dos termos. O problema atual é o de como proceder diante dos mais de um milhão de currículos cadastrados na PL e da crescente interdisciplinaridade entre as áreas do conhecimento.

Uma alternativa é a de aplicar técnicas de mensuração de palavras para contabilizar a frequência dos termos mais recorrentes. Identificados, os termos poderiam ser reformatados a partir das recomendações de normalização gramatical, ressaltando-se que, por se trabalhar com linguagem bastante especializada, essa tarefa deve contar com a participação de especialistas da área. Com a lista de termos recomendados, o processo de substituição na base é passível de ser automatizado.

Essas sugestões são de caráter corretivo, destinadas a reduzir inconsistências cuja adequação é relativamente fácil. Considerando-se a

quantidade de currículos cadastrados, é de se esperar – baseando-se nos exemplos vistos – um alto índice de ajustes que favorecerão a PL como um SICT. É evidente que as correções dependerão da boa vontade dos “proprietários” de cada currículo.

#### b) Opções de grafia:

O procedimento costuma envolver situações nas quais o mesmo termo ou expressão apresenta grafias diferentes (geralmente em razão da passagem do tempo). Esta ação pode envolver três aspectos distintos de um SICT: o planejamento, a manutenção e a correção. Para o planejamento, que é uma fase anterior à inserção dos conteúdos, é importante prever mecanismos de orientação àqueles que alimentarão o sistema.

Os aspectos de manutenção e correção são inter-relacionados. A manutenção deve ser feita pelos gerentes dos sistemas de informação, que, com o auxílio de especialistas das áreas, podem atualizar a lista de termos no que se refere às opções de grafia. Com a lista atualizada, recomenda-se o uso de remissivas que orientem os usuários na escolha do termo. Com as remissivas, os novos registros que porventura fossem utilizar termos em desuso serão orientados a adotar o termo preferido pelo sistema. No caso dos registros anteriores à atualização de determinados termos, será preciso estabelecer uma rotina que identifique os currículos com tais registros e que recomende ao usuário a correção necessária.

Para a PL, a mesma estratégia sugerida na normalização gramatical também seria válida para opções de grafia, mas o trabalho provavelmente seria maior, pois, além da necessidade de identificação dos termos mais usados na base do sistema, seria preciso analisá-los conforme as áreas de conhecimento, para identificar quais opções de grafia deveriam prevalecer.

#### c) Controle de sinonímia:

A sinonímia é uma relação de equivalência entre, ao menos, duas palavras. Para um sistema de informação interessa o quanto um termo é preferencial para ser utilizado no sistema com relação a outros termos. O ideal é representar o conceito através de um único termo e, assim, permitir a combinação entre a linguagem do usuário e a do sistema. O uso de muitos termos dificulta a compatibilização entre uma estratégia de busca e as formas de representações lingüísticas na base, além de provocar uma dispersão de informações devido ao uso de vários termos para um mesmo conceito.

Um recurso que permita ao sistema “compreender” que dois ou mais termos diferentes tenham o mesmo significado não é trivial. É requerido um grau de especialidade relativamente alto para prever relações de equivalência entre termos e/ou expressões, ou seja, é preciso dominar a

respectiva área do conhecimento. Acrescenta-se que as relações devem seguir um pressuposto nocional capaz de interpretar o significado para determinado domínio ou área de conhecimento para a qual as equivalências devem e podem ser estabelecidas.

Para um sistema como a PL, essa é uma missão bastante penosa – ou até impossível - pois exige a capacidade de lidar com todas as áreas de conhecimento. O fato de a PL abarcar domínios de conhecimento de toda a C&T implica em montar esquemas de relações para cada um dos domínios. Criar um esquema único, capaz de associar universos tão diferentes do conhecimento, seria uma tarefa extremamente complexa – quiçá impossível, em todo caso fadada ao insucesso - dada a multiplicidade de universos semânticos.

Outro caso importante no âmbito da C&T é a mudança de termos decorrentes da consolidação da terminologia da área: não são raros os casos de uso de um termo que, com o passar do tempo, cai em desuso ou então se transforma em um outro, que passa a vigorar na linguagem da área. Termos como “MEIO AMBIENTE”, “AIDS” ou “PORTADOR DE NECESSIDADES ESPECIAIS”, por razões diferentes, foram cunhados recentemente, fruto de processos sociais que sempre estarão presentes no ambiente da C&T.

#### d) Controle de Homonímia:

A homonímia é o fenômeno pelo qual diferentes entidades são designadas pela mesma palavra. Ela ocorre entre itens com significados diferentes que possuem o mesmo som e a mesma grafia (homônimos perfeitos como “literatura” - substantivo e “literatura” - disciplina), ou o mesmo som (homônimos homófonos como caça - ato de caçar e cassa - tornar sem efeito), ou apenas a mesma grafia (homônimos homógrafos como o verbo “seco” e o adjetivo “seco”).

Na PL a homonímia torna-se um problema muito mais grave, em razão de dois fatores. Primeiro, o SRI não é capaz de diferenciar as mais simples relações sintáticas; ou seja, numa procura pelo termo porta são recuperáveis todos os currículos nos quais a palavra PORTA está presente e ainda os currículos com a palavra PORTA como radical, por exemplo: PORTA-enxertos, comPORTamento, imPORTAção. Segundo, além do problema com os radicais, o sistema não evita palavras irrelevantes para os processos de recuperação da informação, as chamadas *STOPWORDS*, geralmente compostas de preposições, artigos ou conjunções<sup>2</sup>. Em tal ocorrência, se for feita uma busca por PARÁ (estado brasileiro), o resultado considerará todos os currículos que contenham a preposição PARA.

Inicialmente, é urgente a necessidade de filtragem das *Stopwords* no sistema de recuperação da PL. Nesse caso, é preciso criar uma lista de

<sup>2</sup> Em alguns casos, como nos termos compostos, o uso de preposição dá significado ao termo, por exemplo: Tecnologias de Informação. Neste caso trata-se de sintagmas, que devem ter um tratamento diferenciado a partir de sua identificação

termos indesejáveis (excetuados os casos em que os mesmos compõem sintagmas), e tais termos devem ser desconsiderados pelo sistema quando o mesmo gerar a lista de índices. Este é um procedimento interno que não envolve o preenchimento dos currículos, pois não há como sugerir que os usuários evitem preposições, artigos, conjunções, advérbios e outras palavras comumente consideradas *stopwords*, pois se tornaria inviável o preenchimento de campos que utilizam a linguagem natural, tais como o campo TÍTULO.

Os termos compostos (também denominados sintagmas) são freqüentes em domínios especializados e podem ser formados pela soma de dois termos ou, até mesmo, por uma construção sintagmática mais complexa. Assim, os termos compostos são formados por palavras ou por radicais que pertencem a classes de palavras diversas. A seguir, são enumerados alguns sintagmas que foram extraídos a partir das palavras-chave dos exemplos analisados na SciELO: Coelho Doméstico, Letramento Digital, Mecânica Estatística, Membrana de Barreira, Campos Cristalinos, Saúde Coletiva, Impacto Bibliográfico, Mecanismos de Busca, Cenários Futuros, Tempos de Relaxação.

#### e) Organização dos termos:

Os desenvolvedores da PL, a partir da Tabela de Áreas do Conhecimento do CNPq, criaram um recurso que conduz o usuário na escolha, dentro de um plano classificatório, de área(s) do conhecimento referentes às suas produções bibliográficas, técnicas, ou artísticas/culturais.

Essa classificação das Áreas do Conhecimento, usada pela PL, segue uma estrutura arborescente similar a um plano de classificação, que é um tipo de vocabulário controlado. Ambos – a classificação da PL e um plano de classificação – têm por base o princípio da hierarquia, que oferece como vantagem o fato de, ao ordenar as atividades hierarquicamente, possibilitar uma visão do conjunto e de como essas se distribuem. A desvantagem está na necessidade de se ampliar o universo de escopo com níveis mais complexos.

Do ponto de vista da organização da informação, esses dois recursos da PL partem do pressuposto de que tanto as Áreas de Conhecimento quanto os Setores de Aplicação e suas respectivas subdivisões organizam-se em classes auto-excludentes. Tal pressuposto, face ao disposto pela Teoria da Classificação desenvolvida na área da Biblioteconomia, é correto. Ressalta-se que o sistema não impede que o usuário cadastre mais de uma opção ou acrescente informações no nível mais específico. A inclusão de novas opções pelo usuário, no entanto, abre a possibilidade da inclusão de sinônimos ou de uma classe que não seja auto-excludente em relação aos termos já previstos pelo sistema.

Numa primeira visão, é possível entender que, para um sistema voltado a um contexto informacional tão amplo como a PL, seriam



necessários (mesmo que somente para indicar as Áreas de Conhecimento e os Setores de Aplicação) níveis de especificidade mais aprofundados ou talvez mais categorias em cada nível. Entretanto, é justificável a opção generalista e reducionista dos projetistas do sistema: os dados coletados nestes campos são utilizados para fins de produção de indicadores e, como se sabe, quanto mais dispersos e fragmentados, menor será a possibilidade de se estabelecerem agrupamentos homogêneos, suficientemente capazes de demonstrar algum comportamento da sociedade científica ou tecnológica.

Por outro lado, existe também a possibilidade do preenchimento com termos generalizantes ser percebido como desestimulante, em razão dos pesquisadores encontrarem dificuldades para relacionar suas produções com termos generalistas oferecidos pelo sistema. Um sinal dessa situação foi percebido nos exemplos examinados neste estudo. Verificou-se que o campo Setores de Atividades é pouco preenchido e o de Áreas do Conhecimento apresenta, freqüentemente, diferentes escolhas entre os autores de um mesmo artigo. Para conclusões mais precisas, uma análise mais criteriosa – orientada a procedimentos estatísticos de amostragem – seria necessária: fica aqui o registro da sugestão para futuras pesquisas.

Se uma análise da base de currículos da PL conseguir responder às necessidades para as quais o sistema foi desenvolvido, entende-se que uma mudança não é prioritária; caso contrário, uma avaliação da classificação das Áreas de Conhecimento e Setores de Aplicação será importante, visando uma provável reformulação. Tal ação é indicada considerando-se que o problema aumentará numa razão proporcional à inclusão de mais pesquisadores e também ao aumento da produção científica, técnica e artística nacional.

Caso se comprove a necessidade de reformulação da classificação, sugere-se observar o processo de construção de um plano de classificação<sup>33</sup> que incorpore o controle de vocabulário. Tal tarefa exige a composição de equipes formadas por especialistas nas respectivas áreas de conhecimento e também de pessoal capacitado para elaborar ferramentas orientadas à organização da informação. O processo se desdobra nas seguintes etapas:

1. levantamento das listas livres (no caso da PL, uma para Áreas de Conhecimento e outra para Setores de Atividades);
2. análise crítica, se necessário, dos termos incluídos nas listas (verificar sinonímias, consistência em termos de normalização gramatical, opções de grafia e solução dada aos termos compostos);
3. elaboração de listas alfabéticas consistentes de áreas e setores, desdobradas em suas respectivas especificidades, caso preciso, e acrescidas das remissivas que se fizerem necessárias;

<sup>3</sup> A sugestão de procedimentos segue os passos descritos por Smit e Kobashi (2003).

4. categorização, em maior ou menor grau, dos itens constantes da lista. Nomeação das categorias maiores, novamente incorporando na preocupação o controle de vocabulário;
5. análise das nomeações que podem gerar leituras diferentes e elaboração de notas de escopo ou notas de uso. Tanto as notas de escopo como as de uso serviriam como recursos de orientação para o preenchimento da PL. A elaboração destas notas (sobretudo as de escopo) requer a participação efetiva dos especialistas da área para atribuir, a partir do domínio de conhecimento específico, o conceito próprio ao termo. Os especialistas da área da informação seriam necessários para orientar sobre a importância, a função e, principalmente, a elaboração das referidas notas;
6. submissão das listas (modalidade, categorizada e alfabética) a testes, avaliação do resultado dos testes, incorporação de ajustes e efetiva implantação do plano de classificação.

Outra possibilidade de organização das Áreas de Conhecimento e Setores de Aplicação é proposta pelo tesouro. O tesouro e o plano de classificação são instrumentos para organização da informação que incluem o controle terminológico em graus diferenciados, e são utilizados em sistemas de informação visando traduzir a linguagem dos documentos, dos indexadores e dos pesquisadores numa linguagem controlada, para uso na indexação e recuperação de informações.

Os tesouros apresentam maior flexibilidade na sua elaboração, pois não partem do princípio de uma única hierarquia para organizar os termos. No entanto, para os propósitos desta discussão reforçamos a necessidade da categorização dos termos, quer seja no contexto de um plano de classificação ou de um tesouro. Um aspecto mais recente sobre os tesouros relaciona-os às ontologias. De fato, há entre os tesouros e as ontologias algo em comum: ambos configuram um sistema de conceitos, porém, conforme estudo de Moreira, Alvarenga e Oliveira (2004), os tesouros servem de instrumento de registro e controle terminológico, para uso humano, ao passo que as ontologias objetivam o registro do conhecimento para inferências computacionais.

A ontologia da PL é uma detalhada estrutura das partes que compõem o currículo, representando associações e níveis de subordinação/equivalência entre essas partes. Nesta estrutura há classes e categorias, com seus respectivos atributos. A relação ocorre entre classes e não entre conceitos, ou seja, não existe uma relação nocional, pois não há significado semântico sob o ponto de vista humano. Existe, sim, um sentido dado ao currículo entre as partes que o compõem. Na prática, essas relações, por si só, exprimem somente esquemas de relações genéricas e relações partitivas. Essas relações serão de fato utilizadas quando o sistema de informação que adotar uma determinada ontologia iniciar a inserção de dados em sua base. Na PL, a ontologia tem uso efetivo a partir dos currículos cadastrados, e é a partir deles que

podem ser executadas as inferências computacionais. O processamento automático das inferências permitirá classificar conceitos dentro de uma hierarquia e ainda verificar se determinadas instâncias pertencem a determinadas classes. Exemplo: se existe registrada como título de um livro a frase CONHECIMENTO PÚBLICO, a ontologia já terá previsto que aquele campo específico sempre pertencerá à instância de uma produção bibliográfica específica (livro) e sempre “significará” o título deste tipo de publicação.

A partir da ontologia são elaborados modelos lógicos para verificar inferências conforme as mais diversas finalidades. Um exemplo: deseja-se saber se os bolsistas de produtividade, nos últimos três anos, concentraram suas produções bibliográficas em periódicos da Qualis da Capes. Seria possível, através dos currículos dos pesquisadores, associarem as classes referentes à produção bibliográfica a um sistema externo que, neste exemplo, é a Base Qualis da Capes. O modelo buscaria relacionar as classes e as categorias atinentes às instâncias concernentes à produção bibliográfica de artigos. Para que isso fosse possível, deveria haver compatibilidade entre as ontologias da PL e da Base Qualis.

Uma das deficiências das ontologias, se comparadas aos tesouros, está na dificuldade para criar relações semânticas direcionadas a representações de conceitos. Na PL, por exemplo, a ontologia não altera a representação feita pelo autor através de palavras-chave de um artigo por ele publicado. A ontologia poderá fazer uso do que foi preenchido, mas não ajudará a preencher. Então, ela – a ontologia – funciona na PL como uma meta-estrutura que pode viabilizar relações semânticas, mas não as realiza. Um modelo lógico seria capaz de inferir que nos dois sistemas (tesauro e ontologia) há produções bibliográficas da área da Ciência da Informação, da classe artigo, publicados no ano de 2002. Do ponto de vista semântico, os dois artigos assemelham-se por discutirem temáticas semelhantes, mas este entendimento exigiria uma compreensão de conceitos a partir do domínio da área da Ciência da Informação, e a ontologia não contempla essa interpretação de significados humanos.

Considera-se que o uso híbrido – tesauro e ontologia – seria de grande valia para os SICT, proporcionando, entre outros benefícios, a possibilidade de busca orientada através de disponibilização de tesauro na etapa de definição dos termos para busca e inferências entre SICT distintos, desde que compartilhem de ontologias comuns. O uso híbrido para a organização da informação não é uma sugestão recente, já que no artigo de E.W. Dias (2001) recomenda-se o uso combinado de instrumentos desenvolvidos especificamente para o contexto digital com recursos que já utilizados para fins de organização da informação antes da adoção das tecnologias eletrônicas.

Observou-se que a padronização dos vocabulários, segmentando-os por áreas específicas de conhecimento, possibilitaria a identificação mais adequada de termos adotando-se a forma no substantivo, masculino e singular dos termos. Idealmente, a normalização gramatical em repositórios abertos deve prever ações importantes na etapa inicial do

desenvolvimento do sistema. Investir na elaboração prévia de uma lista de termos especializados é um bom caminho porque, além de tudo, é importante que a SICT desenvolva atividades compartilhadas para o controle de vocabulários.

Uma das discussões atuais sobre os conteúdos da *Internet* volta-se para a utilização compartilhada de recursos entre sistemas disponibilizados na grande rede, discussão essa relacionada com os princípios de *Web Semântica*. Desta maneira, a utilização de vocabulários controlados por parte dos SICT nacionais deve ser estimulada entre os atores que gerenciam tais sistemas, para averiguarem a viabilidade de implantações conjuntas. Reconhece-se que a compatibilização semântica entre os sistemas não é simples, mas a adoção, mesmo que simplificada, de vocabulários controlados em sistemas com pouco ou nenhum controle pode ser benéfica no que diz respeito à organização da informação.

Entende-se que caberia ao planejamento de um SICT da dimensão da PL a criação de contextos segmentados de organização da informação. Na prática, seria necessário elaborar estruturas significantes de termos para domínios específicos de conhecimento. O problema maior reside em pensar nesse aspecto como uma ação corretiva, quando idealmente deveria ser uma atividade da fase de idealização/planejamento do sistema.

Por fim, sugere-se o uso combinado de um tesouro e da ontologia já estabelecida para a PL. Cada um desses recursos tem funções próprias, que não se anulam e tampouco se sobrepõem, mas se complementam. A ontologia não é capaz de favorecer as representações conceituais em forma de palavras, coisa que o tesouro é reconhecidamente capaz de fazer, podendo ser usado tanto pelos usuários que preenchem o sistema, como pelos que buscam informações nele. Ao tesouro não cabe a tarefa de criar uma estrutura de relações que possam produzir inferências lógicas entre suas partes, esse é o papel da ontologia. A ontologia ainda permite o compartilhamento de suas estruturas com outros sistemas, e espera-se que os outros SICT nacionais levem em consideração tal função da ontologia.

## 6 Conclusão

Como síntese das considerações a respeito da PL, inicialmente ressalta-se que o currículo gerado pelo sistema é demasiadamente longo, proporcionando uma leitura confusa e descontextualizada das atividades desenvolvidas pelo pesquisador. Numa visão arquivística, isto o torna inadequado como formato de um documento que deveria espelhar a trajetória do pesquisador. Ainda numa visão arquivística, entende-se que o preenchimento do currículo promove a descontextualização de atividades. Por outro lado, é reconhecidamente positivo o fato de cientistas terem se habituado a registrarem seu histórico acadêmico, tornando possível a criação de um grande acervo de currículos de cientistas brasileiros.

Considera-se necessário inserir mecanismos de controle na forma de preenchimento da PL, pois o aumento da comunidade científica brasileira e o respectivo crescimento da produção desta comunidade evidenciam um aumento proporcional de inconsistências. Recursos de normalização gramatical e/ou orientações interativas que direcionem os usuários no preenchimento do currículo podem trazer benefícios a curto prazo por um custo baixo. Considera-se também inconcebível que um sistema da dimensão da PL mantenha falhas elementares, como erros de digitação. Assim, ações corretivas são urgentes.

Historicamente os sistemas (e as políticas) de informação no Brasil foram descontinuados, prevaleceram novas soluções que negligenciavam antigos problemas. Corrigir as inconsistências atuais da PL é bem mais coerente do que aguardar a futura criação de um novo sistema capaz de solucionar as deficiências. Para tanto, é preciso rediscutir a concepção da PL, pensando-a não mais como uma solução integradora de bases e sim como um sistema voltado à gestão e à política de C&T.

Sobre a problemática dos sistemas abertos de informação (sobretudo os de ICT), entende-se que somente tornar acessível a produção científica não favorece o conjunto maior da comunicação científica. A comunicação científica não é um meio, mas um processo composto por produtores, usuários e recursos que regem esse conjunto. É preciso ter clareza quanto à função de cada novo recurso informacional que será disponibilizado para a comunidade, o que requer uma definição de suas finalidades na fase de planejamento. Se a função prevista para um determinado sistema for o armazenamento de arquivos eletrônicos, aspectos de organização são secundários. Porém, se houver a expectativa de que seja um SICT dotado de recursos de recuperação da informação e/ou de que sirva como fonte para elaborar indicadores de C&T, é imprescindível se conhecer as diretrizes necessárias próprias à organização da informação para os devidos fins.

Por fim, o uso de linguagens documentárias e a conseqüente adoção de vocabulários controlados são criticados devido ao custo no processo de organização da informação. Porém, apesar da desvantagem do custo – que é real –, o controle de vocabulário permite alcançar maior consistência e confiabilidade na informação tornada pública e disponível. Diante da influência da racionalidade econômica na formação dos estoques de ICT, ressalta-se que, enquanto recursos mais eficientes (e mais consistentes) não forem desenvolvidos, a referida racionalidade precisa ser refletida. No estado atual dos SICT não cabem mais escolhas excludentes, ou seja, a adoção de um controle rígido ou a permissão de demasiada liberdade ao sistema: preconiza-se a busca por um meio termo que apresente uma boa relação custo/benefício.

## Referências

CATARINO, M. E.; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *DataGramaZero – Revista de*

*Ciência da Informação*, Rio de Janeiro, v. 8, n. 3, jun. 2007. Disponível em: <[http://www.dgz.org.br/jun07/Art\\_04.htm](http://www.dgz.org.br/jun07/Art_04.htm)>. Acesso em: 21 ago. 2007.

DIAS, E. W. Contexto digital e tratamento da informação. *DataGramaZero – Revista de Ciência da Informação*, Rio de Janeiro, v.2, n.5, art. 01, out. 2001. Disponível em: <[http://www.datagramazero.org.br/out01/Art\\_01.htm](http://www.datagramazero.org.br/out01/Art_01.htm)>. Acesso em: 24 jan. 2002.

GRUPO STELLA. *Plataforma Lattes*. Disponível em: <<http://www.stela.ufsc.br/legado/revistaplataformalattes.pdf>>. Acesso em: 18 maio 2007.

KOBASHI, N. Y.; SANTOS, R. N. M. Institucionalização cognitiva da pesquisa científica no Brasil sob a ótica da Ciência da Informação. *Journal of the American Society of Information Science*, 2007. [No prelo].

MARCONDES, C. H. Metadados: descrição e recuperação de informações na web. In: MARCONDES, C. H. et al. *Bibliotecas digitais: saberes e práticas*. 2. ed. Salvador: EDUFBA; Brasília: IBICT, 2006. cap. 2, p. 95-111.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. *Ciência da Informação*, Brasília, v. 3, n. 31, p.42-53, set./dez. 2002.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. P. O nível do conhecimento e os instrumentos de representação: tesouros e ontologia. *DataGramaZero – Revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 6, dez. 2004. Disponível em: <[http://www.dgz.org.br/dez04/Art\\_01.htm](http://www.dgz.org.br/dez04/Art_01.htm)>. Acesso em: 27 mar. 2006.

NORUZI, A. Folksonomies: why do we need controlled vocabulary? *Webology*, v.4, n.2, Jun. 2007. Disponível em: <<http://www.webology.ir/2007/v4n2/editorial12.html>>. Acesso em: 17 ago. 2007.

SMIT, J. W.; KOBASHI, N. Y. *Como elaborar vocabulário controlado para a aplicação em arquivos*. São Paulo: Arquivo do Estado, Imprensa Oficial do Estado de São Paulo, 2003. (Como fazer, 10).