



Instituto Superior de Ciências  
do Trabalho e da Empresa

**Departamento de Ciências e Tecnologias da Informação**

## **Organização da Informação em *Subject Gateways***

**Paulo Jorge de Oliveira Leitão**

**Dissertação para a obtenção do Grau de Mestre em  
Estudos de Informação e Bibliotecas Digitais**

**Orientador: Dr<sup>a</sup>. Fernanda Maria Campos  
Co-Orientador: Prof. Dr. Pedro Faria Lopes**

**Setembro 2004**

## RESUMO

O presente trabalho aborda um tipo de serviço de acesso à Internet designado por *Subject Gateway*. Este tipo de serviços foi criado para servir as necessidades de informação da comunidade universitária e caracteriza-se por seleccionar recursos *Web* de acordo com critérios de qualidade, descrever essa informação com recurso a linguagens de metadados estruturadas, e proporcionar o acesso através de palavra-chave e estruturas temáticas de navegação, com um particular desenvolvimento desta última forma de acesso. Estes serviços são analisados do ponto de vista dos processos e formas de organização da informação.

São analisados especificamente os cinco *Gateways* iniciais (SOSIG, HUMBUL; EEVL, PSIGATE e BIOME) da rede RDN (*Resource Discovery Network*) no que diz respeito às linguagens de metadados utilizadas e às formas de organização da informação, com o objectivo de não só conhecer esses processos de organização, mas também avaliar a sua qualidade em relação às expectativas que o público-alvo tem de um sistema de organização da informação bibliográfico deste tipo. Partindo de um quadro teórico específico que define as características e objectivos de qualquer sistema de informação bibliográfico, bem como as características das linguagens de metainformação, consideradas enquanto instrumentos de operacionalização desses objectivos, a análise desenvolveu-se em duas vertentes: do ponto de vista conceptual, através do estudo de todos os normativos para a descrição utilizados por cada um dos serviços, e do ponto de vista da prática, utilizando uma amostra significativa de descrições de cada serviço e das estruturas de organização da informação para efeitos de *browsing*. Num e noutro caso, a realidade foi analisada e avaliada de acordo com um conjunto de critérios que derivaram do quadro teórico definido.

Verifica-se a aplicação de um modelo baseado nos processos tradicionais de organização da informação, mas com diversas alterações no sentido da simplificação descritiva e da adaptação ao tipo de recursos a descrever, privilegiando claramente a representação e o conseqüente acesso pelo assunto. A qualidade da metainformação sobre o assunto bem como a qualidade global das descrições dos recursos permitem a estes serviços ir ao encontro de algumas das principais expectativas dos utilizadores. As dificuldades mais relevantes dizem respeito à possibilidade de os utilizadores encontrarem todos os documentos de um mesmo Autor, todos os documentos pertencentes à mesma **Obra**, todos os documentos de uma mesma Versão, e, por último, de poderem Navegar no universo descritivo construído.

**Palavras-chave:** *Subject Gateways*, RDN, SOSIG, HUMBUL, PSIGATE, EEVL, BIOME, Metadados, *Browsing*, Organização da Informação.

## **ABSTRACT**

The present dissertation is about Subject Gateways a kind of Internet access service created to serve the academic community's information needs, characterized for selecting web resources accordingly with quality criteria, describing information with structured metadata languages and to give access through keywords and particularly browsing structure's development. These services are analyzed from the point of view of processes and ways of organizing information.

The five initial Gateways (SOSIG, HUMBUL, EEVL, PSIGATE, BIOME) of the RDN (Resource Discovery Network) are specifically analyzed regarding to used metadata languages and ways of organizing information. This analysis aims to know the process of organizing information, and also evaluate their quality in relation to target audience expectation about systems of organizing bibliographic information of this kind. Within a specifically theoretical framework witch defines the characteristics and aims of any bibliographic information system as well there metadata languages, this analyze was developed in two different ways: from the conceptual point of view, through the study of all cataloguing rules of the services, and from the practical point of view, using a significant sample of each services descriptions and their browsing structures. In both cases the services were analyzed and evaluated accordingly with the criteria defined from the proposed theoretical framework

These services use a model based in traditional processes of information organization, with several adjustments in the sense of descriptive simplification and adaptation to the described resources, with special concerns about subject access and representation. The metadata quality about subject as well as the global quality of resources description allows that these services meet some of main user's expectations. The problems are those related with finding all the documents about the same Author, all documents belonging to the same Work, all documents belonging to the same Version, and, at last, to be able to Navigate in the bibliographic universe.

**Keywords:** Subject Gateways, RDN, SOSIG, HUMBUL, PSIGATE, EEVL, BIOME, Metadata, *Browsing*, Information Organization

## **AGRADECIMENTOS**

Este trabalho não teria sido possível sem o conhecimento, o interesse e a paciência da Dr<sup>a</sup> Fernanda Maria Guedes de Campos, a quem agradeço reconhecidamente.

Ao Prof. Pedro Faria Lopes agradeço a disponibilidade sempre presente para acompanhar a elaboração deste trabalho.

Aos meus colegas de Mestrado, particularmente ao Dr. Rui Neves, à Dr<sup>a</sup> Ângela Pereira e à Dr<sup>a</sup> Lurdes Saramago estou grato pela disponibilidade e espírito de colaboração que demonstraram.

Aos meus filhos, por terem tido a paciência de esperar, muitas vezes, pela disponibilidade do pai.

À Alda Ribeiro pela paciência, disponibilidade e aconselhamento nunca regateados.

# ÍNDICE

<b>CAP. I – INTRODUÇÃO</b> .....	<b>8</b>
1.1. TEMA .....	8
1.2. OBJECTIVOS.....	8
1.4. METODOLOGIA GERAL .....	9
1.5. PANORÂMICA DA ESTRUTURA DA DISSERTAÇÃO .....	10
<b>CAP. II – ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO NA WEB E A SOLUÇÃO SUBJECT GATEWAYS</b> .....	<b>11</b>
2.1. CARACTERÍSTICAS GERAIS DA INFORMAÇÃO QUE CIRCULA NA <i>WEB</i> .....	11
2.2. ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO E A SOLUÇÃO <i>SUBJECT GATEWAYS</i> .....	15
2.2.1. <i>Tipologia e características dos serviços de recuperação da informação</i> .....	15
2.2.2. <i>Vantagens e inconvenientes</i> .....	22
2.2.3. <i>Subject Gateways</i> .....	23
2.3. SISTEMAS DE ORGANIZAÇÃO DA INFORMAÇÃO .....	27
2.3.1. <i>Objectivos de um sistema de informação bibliográfica</i> .....	28
2.3.2. <i>Descrição dos recursos de informação</i> .....	29
2.4. METADADOS.....	32
2.5. ANÁLISE E REPRESENTAÇÃO DO CONTEÚDO .....	39
2.5.1. <i>Linguagens de representação</i> .....	39
2.5.2. <i>A utilização das linguagens para a representação do assunto no contexto de metadados</i> .....	41
2.5.3. <i>Browsing de recursos</i> .....	43
2.5.4. <i>Descrição em texto livre</i> .....	46
<b>CAP. III – OS GATEWAYS NO ÂMBITO DA RDN (RESOURCE DISCOVERY NETWORK)....</b>	<b>50</b>
3.1. O NASCIMENTO DA IDEIA E SUAS CONCRETIZAÇÕES .....	50
3.1.1. <i>O projecto eLib e o surgimento dos primeiros Subject Gateways</i> .....	50
3.1.2. <i>A RDN (Resource Discovery Network)</i> .....	51
3.2. PARADIGMAS E INSTRUMENTOS DE ORGANIZAÇÃO DA INFORMAÇÃO PARA <i>SUBJECT GATEWAYS</i> .....	53
3.2.1. <i>O “paradigma” DESIRE (Desire Information Gateways Handbook)</i> .....	53
3.2.2. <i>As Regras de Catalogação ROADS (Resource Organization and Discovery in Subject- Based Services)</i> .....	57
3.3. ORIGEM, EVOLUÇÃO E CARACTERÍSTICAS ACTUAIS DOS <i>GATEWAYS</i> .....	64
3.3.1. <i>HUMBUL: humanities hub</i> .....	64
3.3.2. <i>SOSIG: social science information gateway</i> .....	66
3.3.3. <i>PSIGATE: the physical sciences information gateway</i> .....	68
3.3.4. <i>EEVL: the Internet guide to engineering, mathematics and computing</i> .....	69
3.3.5. <i>BIOME: your guide to Internet resources in the health and life sciences</i> .....	72
<b>CAP. IV – ORGANIZAÇÃO DA INFORMAÇÃO NOS SUBJECT GATEWAYS DA RDN.....</b>	<b>75</b>
4.1. METODOLOGIA DE ANÁLISE DOS PROCESSOS E FORMAS DE ORGANIZAÇÃO DA INFORMAÇÃO .....	75
4.2. LINGUAGENS DE METADADOS .....	80
4.2.1. <i>SOSIG</i> .....	80
4.2.2. <i>PSIGATE</i> .....	85
4.2.3. <i>BIOME</i> .....	89
4.2.4. <i>EEVL</i> .....	91
4.2.5. <i>HUMBUL</i> .....	93
4.3. ANÁLISE E REPRESENTAÇÃO DO CONTEÚDO .....	98
4.3.1. <i>Utilização das linguagens documentais para efeitos de Classificação e Indexação</i> .....	98
4.3.2. <i>Análise das estruturas de browsing</i> .....	106
4.3.2.1. <i>SOSIG</i> .....	106
4.3.2.2. <i>HUMBUL</i> .....	108
4.3.2.3. <i>PSIGATE</i> .....	110
4.3.2.4. <i>BIOME</i> .....	113
4.3.2.5. <i>EEVL</i> .....	119
4.3.3. <i>Descrições em texto livre</i> .....	127
<b>CAP. V – CONCLUSÕES</b> .....	<b>134</b>
<b>BIBLIOGRAFIA</b> .....	<b>139</b>

## ÍNDICE DE TABELAS

TABELA 1 - TIPOLOGIA DOS SERVIÇOS DE PESQUISA NA INTERNET .....	16
TABELA 2 – ATRIBUTOS RELATIVOS À OBRA EM ROADS .....	58
TABELA 3 – ATRIBUTOS RELATIVOS AO DOCUMENTO EM ROADS .....	59
TABELA 4 – ATRIBUTOS RELATIVOS À OBRA NO <i>TEMPLATE RESOURCE</i> .....	60
TABELA 5 – ATRIBUTOS RELATIVOS AO DOCUMENTO NO <i>TEMPLATE RESOURCE</i> .....	61
TABELA 6 – ATRIBUTOS RELATIVOS A RESPONSABILIDADE INTELECTUAL SOBRE O RECURSO NO <i>TEMPLATE RESOURCE</i> .....	62
TABELA 7 – ATRIBUTOS RELATIVOS AO EDITOR NO <i>TEMPLATE RESOURCE</i> .....	62
TABELA 8 – ATRIBUTOS RELATIVOS AO ADMINISTRADOR DO RECURSO NO <i>TEMPLATE RESOURCE</i> ..	62
TABELA 9 – ATRIBUTOS RELATIVOS AO ASSUNTO NO <i>TEMPLATE RESOURCE</i> .....	63
TABELA 10 - GRELHA DAS CATEGORIAS DE ANÁLISE DAS LINGUAGENS DE METADADOS DOS SERVIÇOS EM ESTUDO.....	75
TABELA 11 - INDICADORES DA CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA.....	76
TABELA 12 – CRITÉRIOS E INDICADORES PARA ANÁLISE DA AMOSTRA .....	76
TABELA 13 - ENTIDADES E ATRIBUTOS EM <i>SOSIG</i> .....	145
TABELA 14 - NÍVEL DE CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA EM <i>SOSIG</i> .....	147
TABELA 15 - ENTIDADES E ATRIBUTOS DEFINIDOS EM <i>PSIGATE</i> .....	148
TABELA 16 - NÍVEL DE CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA EM <i>PSIGATE</i> .....	149
TABELA 17 - ENTIDADES E ATRIBUTOS DEFINIDOS EM <i>BIOME</i> .....	150
TABELA 18 - NÍVEL DE CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA EM <i>BIOME</i> .....	152
TABELA 19 - ENTIDADES E ATRIBUTOS EM <i>EEVL</i> .....	153
TABELA 20 - NÍVEL DE CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA EM <i>EEVL</i> .....	153
TABELA 21- ENTIDADES E ATRIBUTOS EM <i>HUMBUL</i> .....	154
TABELA 22- NÍVEL DE CONSECUÇÃO DOS OBJECTIVOS UNIVERSAIS DOS SISTEMAS DE INFORMAÇÃO BIBLIOGRÁFICA EM <i>HUMBUL</i> .....	155
TABELA 23 – ASSUNTOS PRINCIPAIS E SUBDIVISÕES EM <i>SOSIG</i> .....	156
TABELA 24 - DISTRIBUIÇÃO DOS RECURSOS CLASSIFICADOS EM CADA UM DOS NÍVEIS DA ESTRUTURA EM <i>SOSIG</i> .....	157
TABELA 25- DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS NA ESCALA EM <i>SOSIG</i> .....	157
TABELA 26 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS NO INTERVALO 1-5 RECURSOS EM <i>SOSIG</i> .....	157
TABELA 27 - DISTRIBUIÇÃO DOS RECURSOS PELOS TEMAS (SEM ATRIBUIÇÃO MÚLTIPLA) EM <i>HUMBUL</i> .....	158
TABELA 28 – ASSUNTOS PRINCIPAIS E SUBDIVISÕES EM <i>PSIGATE</i> .....	158
TABELA 29 – COMPARAÇÃO DO NÍVEL DE ESPECIFICIDADE DA REPRESENTAÇÃO DOS ASSUNTOS NA ESTRUTURA DE ASSUNTOS E NA CLASSIFICAÇÃO DECIMAL DE DEWEY EM <i>PSIGATE</i> .....	158
TABELA 30 - DISTRIBUIÇÃO DOS RECURSOS CLASSIFICADOS EM CADA UM DOS NÍVEIS DA ESTRUTURA DE <i>BROWSING</i> EM <i>PSIGATE</i> .....	159
TABELA 31 - RECURSOS CLASSIFICADOS EM CADA NÍVEL DE ACORDO COM A ESCALA EM <i>PSIGATE</i> .....	160
TABELA 32 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS NO INTERVALO 1-5 EM <i>PSIGATE</i> .....	160
TABELA 33 – ASSUNTOS PRINCIPAIS E SUBDIVISÕES NO DIRECTÓRIO <i>AGRIFOR</i> .....	160
TABELA 34 - DISTRIBUIÇÃO DOS RECURSOS ATRIBUÍDOS EM CADA UM DOS NÍVEIS EM <i>AGRIFOR</i> .....	161
TABELA 35 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS DE ACORDO COM ESCALA EM <i>AGRIFOR</i> .....	161
TABELA 36 - DISTRIBUIÇÃO DO Nº DE RECURSOS NOS ASSUNTOS NO INTERVALO 1-5 EM <i>AGRIFOR</i> .....	162
TABELA 37 - DISTRIBUIÇÃO DO Nº DE RECURSOS NOS ASSUNTOS NO INTERVALO 1-5 EM <i>AGRIFOR</i> .....	162
TABELA 38 - DISTRIBUIÇÃO DOS RECURSOS POR CADA ASSUNTO EM <i>AGRIFOR</i> .....	162
TABELA 39 - EXEMPLO DA LISTA DE ASSUNTOS BASEADA NA <i>CDD</i> COM A RESPECTIVA NOTAÇÃO EM <i>AGRIFOR</i> .....	162
TABELA 40 – NOTAÇÕES UTILIZADAS POR CLASSE DA <i>CDD</i> EM <i>AGRIFOR</i> .....	163

TABELA 41 - COMPARAÇÃO ENTRE AS TERMOS UTILIZADOS NAS DIVERSAS FORMAS DE <i>BROWSING</i> EM AGRIFOR .....	163
TABELA 42 - ASSUNTOS PRINCIPAIS E SUBDIVISÕES EM NATURAL .....	164
TABELA 43 - DISTRIBUIÇÃO DOS RECURSOS CLASSIFICADOS EM CADA UM DOS NÍVEIS DA ESTRUTURA EM NATURAL .....	166
TABELA 44 - DISTRIBUIÇÃO DOS RECURSOS PELOS GRANDES TEMAS EM NATURAL .....	166
TABELA 45 - DESCRITORES E RESPECTIVAS NOTAÇÕES EM BIORES.....	167
TABELA 46 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS DE ACORDO COM ESCALA EM VETGATE.....	167
TABELA 47 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS EM VETGATE .....	168
TABELA 48 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS DE ACORDO COM ESCALA EM OMNI.....	168
TABELA 49 - EXEMPLO DA UTILIZAÇÃO DE ACM COMPUTING CLASSIFICATION PARA A CLASSE PRINCIPAL "COMPUTING METHODOLOGIES" (3 PRIMEIRAS SUB-CLASSES)EM EEVL.....	168
TABELA 50 - NÍVEIS DE SUBDIVISÃO DOS ASSUNTOS EM EEVL.....	170
TABELA 51 - DISTRIBUIÇÃO DOS RECURSOS CLASSIFICADOS EM CADA UM DOS NÍVEIS DA ESTRUTURA DE <i>BROWSING</i> EM EEVL.....	170
TABELA 52 - NÚMERO DE RECURSOS CLASSIFICADOS EM CADA ASSUNTO EM EEVL.....	170
TABELA 53 - DISTRIBUIÇÃO DOS RECURSOS PELOS ASSUNTOS EM EEVL.....	171
TABELA 54 - ANÁLISE DE CONTEÚDO DO ELEMENTO DESCRIÇÃO NAS REGRAS DE DESCRIÇÃO DOS SERVIÇOS.....	171
TABELA 55 - EXTENSÃO DOS RESUMOS .....	171
TABELA 56 - COMPLEXIDADE DAS PALAVRAS (NÚMERO MÉDIO DOS CARACTERES POR PALAVRA)...	172
TABELA 57 - COMPLEXIDADE DAS FRASES (NÚMERO DE PALAVRAS POR FRASE).....	172
TABELA 58 - FRASES NA VOZ PASSIVA (%) .....	172
TABELA 59 - FLESH READING EASE .....	172
TABELA 60 - RELAÇÃO ENTRE OS RESULTADOS DO TESTE FRE E A EXTENSÃO DOS RESUMOS.....	172
TABELA 61 - RELAÇÃO ENTRE OS RESULTADOS DO TESTE FRE E A EXTENSÃO DA FRASE.....	173
TABELA 62 - RELAÇÃO ENTRE OS RESULTADOS DO TESTE FRE E A COMPLEXIDADE DA FRASE.....	173

# Cap. I – Introdução

## 1.1. Tema

O tema central deste trabalho são os *Subject Gateways*, encarados sob o ponto de vista dos processos e formas de organização da informação que implementam.

*Subject Gateways* são serviços de acesso à Internet, sobretudo à *Web*, desenvolvidos para públicos-alvo específicos, maioritariamente constituídos por especialistas, quer do ponto de vista académico quer profissional, e por aprendizes, quer dizer, a comunidade discente universitária. Baseiam-se numa criteriosa selecção dos recursos *Web* de acordo com as necessidades dos utilizadores e apertados critérios de qualidade; estes são posteriormente descritos de acordo com linguagens de metadados definidas para cada serviço e recuperados essencialmente por duas vias: a pesquisa por palavra-chave na metainformação, e o *browsing* de estruturas temáticas organizadas de acordo com linguagens de assunto que tem origem, maioritariamente, em classificações, tesouros ou listas de cabeçalhos já utilizados pelas bibliotecas.

Existe actualmente um considerável número de implementações deste conceito, quase sempre da responsabilidade de bibliotecas, quer universitárias, como se verifica no caso que analisaremos, quer nacionais, onde se destaca o caso da Biblioteca Nacional da Austrália. A expansão deste tipo de serviços está muito relacionada, na Europa, com o projecto DESIRE<sup>1</sup>, que se desenvolveu na segunda metade dos anos 90 e com o projecto inglês eLib<sup>2</sup> (*Electronic Libraires*) do início da mesma década.

## 1.2. Objectivos

Constituem objectivos deste trabalho:

- a) Analisar as formas de organização da informação desenvolvidas pelos *Subject Gateways* a partir do estudo de um caso paradigmático: os primeiros *Gateways* criados no âmbito da RDN (*Resource Discovery Network*).
- b) Avaliar a eficácia das formas de organização aplicadas na prossecução dos objectivos a atingir por qualquer sistema de informação bibliográfico, objectivos definidos de acordo com as expectativas dos utilizadores.

## 1.3. Motivação para a escolha do tema

Os *Subject Gateways* traduzem, como teremos oportunidade de verificar, a aplicação de um modelo clássico da organização da informação, se bem que adaptado, ao ambiente dos recursos *Web*. A emergência das redes de informação contemporânea e as soluções inicialmente encontradas para assegurar a recuperação da informação nesse contexto foram, e continuam a ser, objecto de discussão e investigação. Inicialmente ausente, de certa forma, destes problemas a tradição biblioteconómica intervém a meio dos anos 90 trazendo para a discussão as vantagens e desvantagens da utilização para este ambiente

---

<sup>1</sup> Disponível em <http://www.desire.org>

<sup>2</sup> Disponível em <http://www.ukoln.ac.uk/services/elib>

do modelo que tão bem conhecia e havia testado ao longo de séculos. Desde os mais radicais, que preconizavam a utilização directa dos instrumentos clássicos, até aos campeões dos metadados, a discussão foi-se desenvolvendo até se perspectivar uma solução que designaremos, para já, de 3ª via que pretende, mantendo as vantagens do modelo clássico de organização da informação, adaptá-lo a um novo ambiente e novos recursos de informação. Assim, pretende-se estudar de que forma esta perspectiva de harmonização se concretizou num tipo de serviço específico de acesso à *Web*.

#### 1.4. Metodologia geral

A opção metodológica base deste trabalho é o estudo de caso. Considerando que o universo dos *Subject Gateways* é, actualmente, extenso não seria possível uma análise exaustiva a todas as implementações deste conceito. Assim, seleccionou-se um caso considerado paradigmático: os primeiros *Subject Gateways* criados no âmbito do que é hoje designado por RDN (*Resource Discovery Network*), que são os seguintes: SOSIG, HUMBUL, EEVL, PSIGATE e BIOME. A escolha deste caso justifica-se porque estes *Gateways* constituem das primeiras e mais desenvolvidas implementações do conceito na Europa. Verifica-se inclusive a sua participação (sobretudo o *Gateway SOSIG, social science information gateway*) na investigação, desenvolvimento e consolidação deste tipo de serviços à escala mundial.

A abordagem deste caso foi precedida da análise da literatura relevante sobre a *Web* que permitisse, por um lado, caracterizar os tipos de informação que nela circulam e os serviços que a permitem recuperar, e sobre organização da informação, quer no que diz respeito ao modelo tradicional quer quanto a metadados. Procedeu-se igualmente a uma revisão dos estudos existentes sobre *Subject Gateways*, bem como a uma análise de conteúdo pormenorizada das construções teórico-práticas específicas para a análise do caso, a saber: as propostas organizativas plasmadas no *Desire Information Gateways Handbook*<sup>3</sup> e as Regras de Catalogação<sup>4</sup> desenvolvidas no âmbito do projecto ROADS (*Resource Organization and Discovery in Subject- Based Services*)

A partir da análise da literatura, é definido um conjunto de critérios de análise / avaliação aplicados em seguida ao estudo do caso seleccionado. Estes critérios encontram-se desenvolvidos no ponto 4.1. e são definidos, quando tal se aplica, para as duas vertentes possíveis de análise da organização da informação: do ponto de vista das concepções, neste caso, dos normativos definidos para cada serviço, e do ponto de vista da prática, aspecto analisado a partir de uma amostra de todos os serviços objecto de análise.

Os aspectos relativos a funcionalidades de pesquisa só serão abordados de forma operacional, ou seja, na medida em que servem para testar a aplicação de algumas das formas de organização da informação. Por outro lado, e dado que a perspectiva de abordagem dos processos de organização da informação é a dos resultados para o utilizador, excluíram-se da análise todos os aspectos relativos aos metadados de administração das descrições.

---

<sup>3</sup> Disponível em <http://www.desire.org/hanbook>

<sup>4</sup> ROADS cataloging guidelines. Disponível em <http://www.ukoln.ac.uk/metadata/roads/cataloguing/cataloguing-rules.html>

## 1.5. Panorâmica da estrutura da dissertação

No segundo capítulo serão abordadas as questões de contextualização global destes serviços, com uma breve panorâmica sobre a informação que circula na *Web* e sobre os diversos serviços que a pretendem recuperar, entre os quais os *Subject Gateways* (Pontos 2.1 e 2.2). Em seguida, analisam-se as características dos sistemas de informação bibliográfica e dos tipos de linguagens da descrição, incluindo as linguagens de metadados (Pontos 2.3 e 2.4). Considerando que os atributos que têm como objectivo representar o(s) assunto(s) contidos nos objectos de informação assumem um papel de destaque, analisam-se teoricamente os vários processos e formas dessa representação (Ponto 2.5).

A caracterização do caso em análise – os *Gateways* da RDN – é objecto de tratamento no Capítulo 3º, onde se descreve a origem e evolução dos serviços a estudar (Ponto 3.3), contextualizando-os no seu ambiente específico, quer organizacional (Ponto 3.1), quer conceptual e operacional (Ponto 3.2).

No Capítulo 4º procede-se a uma avaliação extensa dos processos de organização da informação nos cinco serviços seleccionados, analisando as linguagens de metadados utilizadas (Ponto 4.2), destacando em ponto autónomo os aspectos que dizem respeito à representação do conteúdo, a saber: classificação e indexação (Ponto 4.3.1), estruturas de *browsing* (Ponto 4.3.2) e descrições em texto livre (Ponto 4.3.3). Desta análise resultou um conjunto de tabelas que, por questões de legibilidade do texto, foram reunidas na parte final da dissertação.

## Cap. II – Organização e recuperação da informação na *Web* e a solução *Subject Gateways*

### 2.1. Características gerais da Informação que circula na *Web*

Constitui uma evidência indiscutível que a *Web* é, senão o maior, pelo menos um dos maiores repositórios de informação do mundo contemporâneo. Mas, ao contrário do que acontece com a maioria dos outros tipos de informação detidos pelas bibliotecas, a que circula neste media apresenta características diferenciadas que importa referenciar, na medida em que constituem aspectos que condicionam as formas de organização.

Esta especificidade começa, desde logo, a relevar-se no que diz respeito à produção da informação e desenvolve-se nas fases seguintes de disponibilização e manutenção. A natureza do media e a utilização social que dele é feita condiciona igualmente o tipo de conteúdos disponibilizados, apresentando-se com grande diversidade e variedade de tipologias documentais.

Os aspectos relativos à produção, manutenção e difusão podem designar-se de estruturais na medida em que caracterizam basicamente qualquer tipo de conteúdo acessível através deste media e dessa forma serão tratadas neste trabalho. A área dos conteúdos apresenta características de maior individualidade, embora seja possível, como veremos, apontar algumas tendências.

O modelo de produção de informação para a *Web* apresenta uma característica básica que é a da hipertextualidade, ou seja, a possibilidade de estabelecer ligações entre partes da informação de forma simples, adaptável a uma grande variedade de circunstâncias e de fácil utilização. Não só os documentos produzidos, nomeadamente os sítios, se baseiam nesta possibilidade, como é a partir dela que se tece a teia de relações entre os diversos conteúdos independentes. Na sua máxima utilização, esta possibilidade permite a produção de documentos de carácter absolutamente secundário, ou seja, colecções mais ou menos organizadas de ligações a outros sítios e documentos. Do ponto de vista interno um Sítio é sempre um conjunto de pedaços de informação relacionados através de múltiplas ligações. No entanto, a utilização de ligações, nomeadamente externas, nem sempre constitui um valor acrescentado, podendo ao contrário revelar-se como um factor de complexidade acrescida na apropriação da informação. Matthew Ciolek define, de forma sintetizada, mas identificando claramente os problemas, esta questão ao escrever: “Para além do mais, as ligações a outros recursos *Web* tendem a ser estabelecidas promiscuamente, isto é sem muita consideração sobre a relevância ou qualidade do recurso alvo. Existe também uma pronunciada circularidade das ligações. (...) Finalmente, a ênfase continua a ser posta no estabelecimento de tantas ligações quanto as possíveis – como se a reputação e a utilidade de um dado recurso online dependesse somente do número de recursos *Web* que cita.” (CIOLEK:1996).

O suporte digital e esta facilidade do estabelecimento de ligações proporcionam a produção de informação que associa vários media resultando no nascimento de documentos multimédia ou, mais apropriadamente designados neste contexto, hipermédia. Esta associação configura o carácter integrador do media e constitui um outro aspecto claramente diferenciador de informação produzida em outros suportes.

A edição de informação neste ambiente é completamente diferente do que era possível até há bem pouco tempo e pode sintetizar-se nos seguintes traços:

- Facilidade de edição individual. Qualquer indivíduo, desde que com alguns conhecimentos técnicos, pode produzir informação e disponibilizá-la online.
- Grandes quantidades de informação não foram objecto de qualquer processo de revisão, crítica e validação
- Qualquer indivíduo ou instituição pode editar um Sítio sobre qualquer assunto, independentemente de estar ou não autorizado a fazê-lo
- Uma grande maioria de produtores de informação não se sente responsável por garantir a continuidade e actualização da informação
- Uma grande maioria de produtores não sente necessidade de informar os possíveis utilizadores sobre as características da informação nem sobre si próprio enquanto produtor.

A questão da manutenção da informação deve ser perspectivada sob vários pontos de vista, quer tecnológico, quer de conteúdo. Neste contexto abordaremos apenas as questões de natureza tecnológica / estrutural, outros aspectos serão abordados aquando da análise das características do conteúdo.

A probabilidade de alteração da localização da informação com a correspondente mudança de *URL (uniforme resource locator)* apresenta-se como um dos problemas mais comuns e que nem sempre é acautelado pelos produtores. A experiência mostra que, enquanto determinados produtores de informação mantêm, pelo menos durante determinado tempo, o anterior URL activo chamando a atenção para a mudança e mesmo redireccionando o utilizador para o novo endereço, muitos há que não partilham destas preocupações impedindo, assim, que se concretize um dos desideratos fundamentais de todo o trabalho de referenciação desenvolvido por uma qualquer biblioteca, qual seja a de garantir ao utilizador a fiabilidade de acesso ao longo do tempo de um determinado documento/informação.

Um outro aspecto, relacionado com o anterior, é o problema das ligações desactivadas, ou seja, a circunstância, tanta vez verificada, de encontrarmos ligações que não conduzem a lado nenhum a partir da página que está a ser consultada. A origem deste fenómeno pode ser variada: a já referenciada alteração de URL, a desactivação do Sítio, problemas técnicos de acesso. É, actualmente possível, através de aplicações que fazem verificação periódica de ligações (*link checkers*) obviar em certa medida a este problema.

A alteração da estrutura do Sítio, em virtude da adaptação de uma nova tecnológica ou de uma reestruturação da sua organização, são aspectos importantes a considerar na medida em que podem condicionar o acesso à informação.

Verificam-se, assim, diferentes probabilidades de alteração de características de tipo estrutural na vida útil dos documentos produzidos para este ambiente, o que os distancia claramente dos que têm um suporte físico estático e para os quais as bibliotecas desenvolveram ao longo de séculos processos de apropriação e organização.

A informação produzida com estas características e neste contexto destina-se a ser difundida da forma mais alargada possível, nunca antes experimentada pelas sociedades humanas, graças a um conjunto básico de tecnologias, mas também a processos sociais,

que possibilitam e incentivam a transferência de informação em moldes mais socialmente alargados.

Uma vez recuperada a informação pretendida, o utilizador depara-se com diversas possibilidades de obter o acesso aos conteúdos, que basicamente podemos agrupar em três categorias:

- ✓ Acesso livre: o utilizador não necessita de se identificar nem de proceder ao pagamento de qualquer contrapartida
- ✓ Acesso pago: o utilizador necessita de proceder ao pagamento de uma determinada quantia
- ✓ Acesso condicionado: o utilizador necessita de pertencer a uma determinada comunidade de utilizadores para obter acesso. Estão neste caso, por exemplo, alguns serviços disponibilizados via *Web* pelas bibliotecas que só podem ser acedidos se o indivíduo fizer parte da comunidade de utilizadores da biblioteca estando devidamente registado nessa comunidade.

Nos últimos anos tem-se assistido a uma tendência para o crescimento de informação disponibilizada contra pagamento. O modelo de negócio presente baseia-se no acesso livre a partes e/ou sínteses dos conteúdos e ao pagamento contra um acesso total. Estão neste caso, por exemplo, muitas das revistas electrónicas que disponibilizam acesso livre aos resumos dos artigos mas não ao texto completo. Apesar desta tendência, existe ainda um grande conjunto de conteúdos de acesso absolutamente livre procedentes de várias origens: pessoais, institucionais e, claro, comerciais.

Uma das primeiras características mais evidentes do universo de conteúdos disponibilizados na *Web* é a da diversidade dos assuntos representados a par da diversidade linguística, embora com um grande peso para os conteúdos em língua inglesa. Esta diversidade temática vai a par com um diferente nível de tratamento dos assuntos, quer em termos de profundidade de abordagem quer de objectivos que se querem atingir ao disponibilizar um determinado tipo de informação. Estão ainda na infância os estudos sobre este aspecto, nomeadamente quanto à representatividade dos vários tipos de conteúdo. No entanto, de acordo com alguns autores podemos referir que: “Uma estimativa generosa pode ser a de que 5% dos recursos disponíveis na *Web* têm valor permanente para o conhecimento e devem especialmente ser armazenados, catalogados e preservados. A grande parte remanescente é maioritariamente informação comercial, muitas vezes com uma orientação publicitária “ (McCALLUM: 2000) A concretizarem-se estas previsões, grande parte dos conteúdos acessíveis têm objectivos que não se enquadram nos critérios de selecção da grande maioria das bibliotecas, bem como uma característica de grande volatilidade que não garante consistência no acesso a longo prazo.

A percentagem dos recursos informativos que valeria a pena referenciar tem, no entanto, e no que diz respeito aos conteúdos diferenças substanciais, quer quanto à profundidade e exaustividade do tratamento dos assuntos, quer quanto à validade e perspectiva de abordagem. Apenas para referir um exemplo, um Sítio com origem numa determinada instituição académica tem, à partida, uma muito mais alta probabilidade de conter conteúdos válidos do que um Sítio sobre o mesmo assunto mas com origem individual, sobretudo se não for possível avaliar as credenciais do autor.

Por outro lado, a informação com objectivos de tipo comercial não tem apenas uma área própria que possa ser claramente identificada e eventualmente excluída. Informação com este tipo de propósitos pode estar incluída em conjuntos que apresentam qualidade e que são avaliados como sendo pertinentes no contexto de uma dada política de selecção.

Embora não existam estudos que permitam tirar conclusões seguras, pensamos que a informação disponível apresenta uma tendência para privilegiar os “acontecimentos” científicos, sociais, culturais mais recentes e manter-se apenas enquanto esses conteúdos apresentam alguma validade imediata. Isto acontece, pensamos, já porque a *Web* é o actual media com maior potencialidade de divulgação mundial, já porque o utilizador espera que seja na *Web* que encontre a “última novidade”, já porque, e este aspecto é claramente muito discutível, este não é o media mais apropriado para acolher um desenvolvimento e exaustividade de nível superior no tratamento dos assuntos, se bem que o seja para efeitos de disponibilização desses conteúdos. Exemplarmente, estamos a referir o caso de uma tese de doutoramento sobre um dado assunto que não é, normalmente, disponibilizada com base nas tecnologias usadas para a produção de documentos hipertextuais (documentos típicos deste media), mas em outros formatos apropriados mais para a transferência do que para leitura e apropriação.

De tudo isto resulta um facto que parece claro: a pertinência de um conteúdo é tanto mais importante quanto esse conteúdo se alterar com relativa rapidez. A sucessiva actualização da informação é, desta forma, uma realidade esperável em praticamente todos os sítios activos, embora, como é evidente, a taxa de actualização seja diferente consoante a natureza do conteúdo. Esta realidade põe vários problemas, entre os quais destacamos: obter informação sobre a actualização; como assegurar que o “velho” conteúdo, que pode continuar a ser pertinente para uma determinada comunidade de utilizadores, continua activo; até que ponto as alterações do conteúdo são de tal forma profundas que inviabilizam a sua inclusão na colecção da biblioteca de acordo com os critérios de selecção e avaliação definidos, que alterações de conteúdo justificam a consideração de uma nova entidade informativa.

Ainda neste contexto, um último aspecto importa referir e que é o da tendência para publicar informação em diferentes estádios de “finalização”. Ou seja, com alguma frequência, vemos, por exemplo, sítios que tem partes mais ou menos substanciais sem qualquer conteúdo, as famosas partes em construção.

Embora a diversidade seja, como vimos, a pedra de toque de todo o universo, é possível agrupar tipologicamente determinados recursos com a vantagem de a categorização poder tornar a análise mais eficaz e para os inexperientes informá-los sobre o que podem encontrar. Uma tipologia dos recursos *Web* poderia ser a que a seguir se apresenta (KOVACS:2000):

- ✓ Motores de pesquisa
- ✓ Directórios de assunto
- ✓ *Subject Gateways*
- ✓ Portais
- ✓ Guias temáticos
- ✓ Bibliotecas electrónicas, virtuais, digitais
- ✓ Associações e organizações
- ✓ Publicações periódicas em formato electrónico

- ✓ Listas electrónicas ou conferências
- ✓ Textos electrónicos
- ✓ Bases de dados disponíveis na *Web*
- ✓ Fontes governamentais
- ✓ Bancos de imagens
- ✓ Produtos e serviços
- ✓ Tecnologias de “*push*” (notícias customizadas, acontecimentos correntes e anúncio de recursos informativos)
- ✓ Leitura de referência
- ✓ Referência (assuntos específicos)
- ✓ Crítica de sítios (assuntos específicos)
- ✓ Bibliotecas especializadas ou departamentos universitários
- ✓ Fontes estatísticas
- ✓ Tutoriais ou cursos de formação

## **2.2. Organização e recuperação da informação e a solução *Subject Gateways***

Existem actualmente diversos serviços que permitem pesquisar partes da *Web*, originárias fundamentalmente de duas fontes: a económica e a institucional não lucrativa. No primeiro caso estes serviços, se disponibilizam de forma gratuita a pesquisa da *Web*, é verdade que o fazem com objectivos económicos claros, donde a existência de todo um conjunto de formas de tornar a informação com objectivos lucrativos muito visível e toda as estratégias de marketing das mais diversificadas. No segundo caso, encontramos serviços que têm origem em instituições oficiais ou como resultado de projectos de investigação, mais preocupados com a qualidade da organização da informação e dos resultados.

### **2.2.1. Tipologia e características dos serviços de recuperação da informação**

Estes serviços implementam perspectivas diferentes sobre a organização da informação a pesquisar e sobre as estratégias de pesquisa que podem ser desenvolvidas. Apesar desta diversidade, é possível definir uma tipologia destes serviços, de acordo com os seguintes critérios: metodologia e critérios de selecção da informação a incluir na base de dados dos serviços, forma de organização da informação, metodologia e estratégias de pesquisa implementadas.

A primeira dificuldade na concretização de uma tipologia é a fixação terminológica. Mesmo na língua inglesa utilizam-se várias designações para este tipo de serviços numa dupla acepção: não só a mesma tipologia de serviço é designada de formas diferentes como a mesma designação é aplicada a tipo de serviços diferenciados.

No entanto, uma análise, ainda que não exaustiva, da literatura produzida sobre este assunto sobretudo com origem no campo dos estudos biblioteconómicos e de informação (THOMAS:2000; CAMPBELL:2001), permite verificar a existência de um esforço de definição tipológica e terminológica num estado que possibilita uma orientação mais clara para a caracterização desta realidade.

Os critérios de classificação definidos nesses estudos podem traduzir-se na seguinte bateria de questões (NICHOLSON: 2000) que especificam os diversos aspectos a ter em conta:

- ✓ Os recursos a incluir na base de dados são seleccionados automaticamente, por humanos ou das duas formas?
- ✓ A base de dados contém recursos com características mais generalistas ou mais específicos?
- ✓ As entradas na base de dados são indexadas automaticamente ou manualmente?
- ✓ Todo o texto é indexado ou apenas partes? Existe uma lista de palavras vazias não incluídas na indexação?
- ✓ O serviço disponibiliza uma árvore de assuntos? Essa árvore de assuntos é construída com base numa linguagem documental ou não?
- ✓ O serviço disponibiliza facilidades de pesquisa por palavra? Nesse caso, que tipo de operadores é possível utilizar
- ✓ Os resumos que acompanham a referência a cada recurso são baseados no contexto em que o termo aparece, são estandardizados ou realizados por humanos?
- ✓ O serviço é um meta-serviço?
- ✓ São utilizadas estruturas de metadados e linguagens documentais para a constituição de um catálogo de recursos?

Submetendo os serviços de pesquisa disponíveis a este conjunto de questões, podemos chegar à seguinte tipologia (Cf. Tabela 1)

**Tabela 1 - Tipologia dos serviços de Pesquisa na Internet**

Designação	Exemplos
Directórios de Assunto (generalistas ou específicos)	Yahoo ( <a href="http://www.yahoo.com">www.yahoo.com</a> ) Open Directory ( <a href="http://www.opendirectory.com">www.opendirectory.com</a> )
Motores de Pesquisa em texto completo	Altavista ( <a href="http://www.altavista.com">www.altavista.com</a> ) Excite ( <a href="http://www.excite.com">www.excite.com</a> ) Infoseek ( <a href="http://www.infoseek.com">www.infoseek.com</a> )
Motores de Pesquisa por extracção	Lycos ( <a href="http://www.lycos.com">www.lycos.com</a> ) Webcrawler ( <a href="http://www.webcrawler.com">www.webcrawler.com</a> ) Hotbot ( <a href="http://www.hotbot.com">www.hotbot.com</a> )
Motores de pesquisa específicos	Fanagalo ( <a href="http://www.fanagalo.co.za">www.fanagalo.co.za</a> )
Portais	IOL ( <a href="http://www.iol.pt">www.iol.pt</a> )
“Subject Gateways”	SOSIG ( <a href="http://www.sosig.ac.uk">www.sosig.ac.uk</a> ) OMNI ( <a href="http://www.omni.ac.uk">www.omni.ac.uk</a> )
Meta Motores	Metacrawler ( <a href="http://www.metacrawler.com">www.metacrawler.com</a> )

Os motores de pesquisa são o tipo de serviço de pesquisa na *Web* mais antigo e baseiam-se no princípio da automação dos processos de selecção e pesquisa de informação. Assim, um serviço deste tipo é normalmente constituído por uma aplicação que, de acordo com princípios e metodologias diferenciadas, pesquisa a *Web* periodicamente e extrai recursos para alimentar a base de dados do serviço. Uma vez descarregados na base de dados, os recursos são indexados automaticamente, também de acordo com diferentes metodologias. Finalmente, existe um interface de pesquisa que permite uma maior ou menor variedade de estratégias de pesquisa dos utilizadores consoante os mecanismos que implementa (operadores booleanos, determinação do peso dos termos numa pesquisa, operadores de proximidade, critérios de restrição, etc.). Os resultados são retornados de acordo com diferentes algoritmos de similitude e

ordenação. Muitos destes serviços implementam, actualmente, algumas possibilidades de alargamento ou restrição dos resultados de pesquisa.

No que diz respeito à cobertura da *Web*, existem dois tipos de motores: os que têm ambições generalistas e os que seleccionam apenas um aspecto específico do universo, normalmente de acordo com uma dada temática.

Para além das diferenças de metodologias de selecção e recuperação da informação que os vários motores de pesquisa implementam, existe uma outra diferença relativa ao método de indexação utilizado, que se revela extremamente importante dado que condiciona em grande parte o tipo de pesquisa que pode ser realizada. Esta diferença conduz à distinção de dois tipos entidades: motores de pesquisa em texto completo e motores de pesquisa por extracção.

Os motores de pesquisa em texto completo indexam todo o texto de todas as páginas *Web* que constam da base de dados. Alguns excluem uma determinada lista de palavras vazias, outros não. Este tipo de motores permite a implementação de uma grande variedade de mecanismos de pesquisa, incluindo a pesquisa de frases completas. A existência de uma lista de palavras vazias pode revelar-se um obstáculo para pesquisas a frases que as incluam maioritariamente, por exemplo “*To be or not to be*”. Parte deste inconveniente encontra-se ultrapassado pela funcionalidade de pesquisa por frases exactas.

Os motores de pesquisa por extracção indexam apenas um número determinado de palavras da página *Web*, que podem ser as palavras mais usadas, um número pré-determinado de palavras ou as palavras do título. Ao contrário do tipo anterior, alguns dos mecanismos de pesquisa e de possibilidades de restrição dos resultados anteriormente referidos não funcionam nestes serviços, por exemplo, a possibilidade de pesquisar frases completas ou realizar pesquisas onde o contexto em que a(s) palavra(s) aparece(m) seja importante.

Os meta motores lançam, a partir de um único interface, uma mesma pesquisa sobre vários motores, apresentando os resultados das várias origens pesquisadas, de forma combinada, de acordo com algoritmos de relevância e de ordenação próprios a cada um dos motores pesquisados ou desenvolvidos pelo próprio meta motor através da avaliação dos resultados retornados em cada um dos serviços individuais.

Os Directórios de Assunto têm como traço organizativo essencial o facto de se basearem numa estrutura de assuntos em árvore, organizada segundo o princípio dedutivo. Esta estrutura pode ter como origem uma determinada linguagem documental ou basear-se num conjunto de assuntos seleccionado pelos autores do serviço como os mais pertinentes e que pode ou não desenvolver-se ao longo do tempo.

A informação indexada em cada um dos assuntos é seleccionada por humanos, bem como os resumos de identificação e caracterização de cada recurso informativo seleccionado.

A pesquisa da informação pode ser realizada, tendencialmente, de duas formas: percorrendo a árvore de assuntos, o que graças à hipertextualidade, permite uma fácil e eficaz navegação entre o geral e o particular; ou através de pesquisa em texto livre que

retorna resultados com base em diferentes algoritmos de similitude e ordenação. O interface de pesquisa em texto livre pode suportar diversas formas de pesquisa (pesquisa simples e/ou pesquisa avançada com utilização de diversos tipos de operadores). Os conteúdos sobre os quais é realizada a pesquisa são normalmente palavras-chave ou vocabulário controlado e os já referidos resumos dos recursos de informação.

Estes directórios podem ter, em termos de assuntos a incluir, ambições generalistas ou específicas concentrando-se sobre um determinado assunto ou área do conhecimento.

Os portais, que em alguns domínios e no ambiente biblioteconómico iniciaram o seu processo de evolução a partir de *Subject Gateways*, são serviços de pesquisa que conjugam um número mais ou menos extenso de serviços ao utilizador, que vão desde o acesso a diversos tipos de recursos informativos, passando pela personalização de interfaces e serviços até a vertentes de comércio electrónico. Como define Debbie Campbell, “Na comunidade biblioteconómica, [portais] podem ser definidos como uma amálgama de serviços ao utilizador onde a conjugação é conseguida através da integração de serviços existentes utilizando agentes de ligação como serviços de customização e autenticação, protocolos de pesquisa como Z39.50, protocolos de empréstimo como ISO 10161 e comércio electrónico. O resultado é um serviço personalizado que permite ao utilizador o acesso a importantes conteúdos quer com origem impressa quer electrónica.” (CAMPBELL: 2001). Em termos gerais, este tipo de serviços oferece, pelo menos, as seguintes facilidades: pesquisa na *Web*, notícias, instrumentos de referência e comércio electrónico.

Se partirmos do princípio de que o problema da recuperação da informação na *Web* é, tal como refere Selberg, “encontrar o conjunto de documentos na *Web* relevantes para a resposta a uma dada pergunta” (SELBERG: 1999) e considerando mesmo que o conceito de relevância é sobretudo relativo ao utilizador e não aos resultados e que pode alterar-se no decurso de uma pesquisa, podemos abordar os vários serviços descritos do ponto de vista do nível e das formas de concretização deste objectivo: encontrar os documentos relevantes para uma pergunta.

Este problema pode ser perspectivado, no entanto, segundo uma óptica quantitativa ou qualitativa, ou seja, trata-se de saber se se pretende encontrar todos os documentos relevantes ou apenas os mais relevantes. De facto, na tipologia de serviços referida existem serviços que apontam para recuperar todos os documentos relevantes para uma pergunta e outros apenas os mais relevantes, pelo que devem ser considerados, em primeiro lugar, de acordo com esta opção.

Os motores de pesquisa são o tipo de serviços que apresentam uma ambição mais alargada em termos quantitativos, levando muitas vezes os utilizadores a assumir que estão a pesquisar toda a *Web*. Ora, de facto, nenhum destes serviços indexa toda a *Web*, o que se deve essencialmente a constrangimentos tecnológicos ligados ao armazenamento de enormes massas de informação e à velocidade de recuperação, proporcionalmente mais baixa à medida do crescimento quantitativo da base de dados. Para além disto, é necessário terem conta outros dois tipos de problemas: o problema da instabilidade dos resultados e o problema de garantir a coerência do índice.

A investigação desenvolvida por Selberg (SELBERG: 1999) demonstra que os resultados apresentam grandes diferenças e que tal não se fica apenas a dever à natureza

essencialmente dinâmica da *Web*, mas ao privilégio dado à velocidade de resposta em vez da qualidade, o que faz com que à medida que a base de dados vai crescendo e a velocidade de resposta é mantida os resultados apresentam sensíveis diferenças.

O mesmo autor demonstra a dificuldade de manter o índice consistente com o conteúdo efectivos dos documentos que são indexados. Passando em revista as duas abordagens possíveis do problema, demonstra, através do seguinte exemplo, a morosidade do processo mais eficaz: “Continuando com o nosso exemplo anterior, se são necessários trinta e quatro dias para indexar um sítio com 50.000 páginas, então são necessários mais trinta e quatro dias para o reindexar completamente. Mesmo usando o método de pesquisa binário para determinar a frequência de extracção, começar ao trigésimo quarto dia significa demorar mais de dois meses para determinar que uma página deste Sítio muda diariamente.” (SELBERG:1999).

Por outro lado, deve ainda considerar-se que dados as diferentes metodologias e princípios utilizados pelos motores de pesquisa para pesquisar a *Web*, identificar e extrair documentos para a sua base de dados nenhum deles indexa o mesmo conjunto de documentos, o que resulta em diferenças, em alguns casos substanciais, de resultados para uma mesma pergunta.

Finalmente, em termos de resultados, os motores de pesquisa apresentam tendencialmente uma elevada taxa de ruído, resultado da ausência de organização da informação. Esta elevada taxa de ruído tem vindo a ser compensada, na generalidade, de três formas: desenvolvendo as opções de pesquisa que se podem utilizar, implementando algoritmos de ordenação por relevância num dado conjunto de resultados, identificando com mais clareza que uma dada página pertence a um determinado Sítio, desenvolvendo métodos contra “spamming”.<sup>5</sup>

O problema com a Internet e os motores de pesquisa para aceder à informação é, assim, como refere Pitschmann “ os resultados de pesquisa podem ser tão vastos que o utilizador não pode esperar avaliá-los num espaço razoável de tempo... Os resultados podem estar muitos afastados, ou ser completamente não relacionados, com os resultados desejados... as tecnologias de inteligência artificial empregues pelos principais serviços de pesquisa são insuficientes para recuperar e avaliar adequadamente conteúdos para ensino e investigação.” (PITSCHMANN: 2001).

Pitschmann refere também que “Pesquisar a *Web* sobre um dado tópico recupera toda a informação pertinente a uma questão, mas não existe qualquer avaliação quantitativa ou forma de filtrar o conteúdo. (PITSCHMANN: 2001). Ele afirma ainda que “Só quando os sítios foram revistos, avaliados, seleccionados e catalogados, os utilizadores serão poupados às ambiguidades resultantes da “quantidade sem qualidade” dos resultados da *Web*. (PITSCHMANN: 2001).

Os meta motores permitem uma mais alta taxa de retorno encontrando mais documentos relevantes para uma pergunta. A investigação de Selberg que temos vindo a referenciar parece indicar que este tipo de serviços, se combinados com formas de incluir os

---

<sup>5</sup> “Spaming” constitui a estratégia utilizada pelos produtores de Sítios Web com o objectivo de aumentar a probabilidade da sua informação ser referenciada pelos motores de pesquisa. A forma mais directa de aumentar esta probabilidade é incluindo nas meta tags de HTML termos normalmente muito pesquisados

resultados de pesquisas anteriores, permite melhores resultados, quer no que diz respeito à identificação de um maior número de recursos permitindo assim ultrapassar as limitações da utilização de um determinado motor, quer a uma mais alta taxa de estabilidade dos resultados ao longo do tempo. No entanto, não podemos deixar de referir que estes serviços, como afirma Scott Nicholson, “combinam resultados de diferentes tipo de bases de dados de páginas *Web* com diversos interfaces, e as referências têm origem numa grande variedade de algoritmos de pesquisa e ordenação. Por causa desta caótica variedade, é difícil executar pesquisas precisas e controladas com estes instrumentos (NICHOLSON:2000)

No outro prato desta balança estão os serviços que não aspiram a uma cobertura em grande extensão da *Web*. Estes serviços desenvolvem-se num largo espectro de possibilidades que vão desde reduzidas exigências de selecção e organização da informação (os directórios de assunto sem utilização de linguagens controladas) até à implementação de políticas bem definidas de selecção e organização da informação (por exemplo, os *Subject Gateways*). Assim, a resposta à questão de quais são os documentos mais relevantes para resposta a uma dada pergunta conhece, sobretudo, diferentes níveis de qualidade no conjunto destes serviços.

Do ponto de vista das facilidades de pesquisa, verifica-se uma tendência para uma certa uniformidade das possibilidades de executar diferentes tipos de pesquisa e que podem ser agrupados em dois grande tipos já referenciados: a pesquisa em texto livre suportando a utilização de diferentes tipos de operadores e a possibilidade de percorrer uma lista de assuntos (*browsing* dos directórios). Praticamente todos os serviços implementam estes dois tipos de pesquisa, mesmo os motores possuem actualmente directórios de assuntos. A diferença é, assim, muito mais ao nível da diversidade em cada um dos tipos (por exemplo, número de operadores que suporta, ou diversidade de hipóteses de percorrer a lista de assuntos) do que de natureza das facilidades.

Do ponto de vista da apresentação dos resultados verificam-se, ao contrário, grandes diferenças. Enquanto que os motores de pesquisa se socorrem apenas de algoritmos para cálculo da relevância de uma página para a resposta a uma dada pergunta e ordenam os resultados de acordo com essa relevância, os outros serviços apresentam formas mais estruturadas de apresentação que passam inclusivé por organizar os resultados por tipologia de recursos, por exemplo, bases de dados, Sítios *Web*, etc. Ainda neste domínio um outro aspecto distingue os dois grandes tipos de serviços e que é a forma de referenciação do recurso apresentado: enquanto que os motores de pesquisa apresentam uma referência que resulta de um automatismo, o que torna frequentemente as descrições quase incompreensíveis, os directórios de assuntos e os *Subject Gateways* apresentam referenciais mais estruturadas com uma qualidade de informação que vai em crescendo dos primeiros para os segundos. A maioria dos *Subject Gateways* apresenta descrições de grande riqueza informativa que incluem palavras-chave extraídas de linguagens documentais ou notações extraídas de classificações.<sup>6</sup>

As diferenças e as similitudes identificadas entroncam em diversas ou semelhantes concepções sobre a questão da recuperação da informação e as funções que os serviços

---

<sup>6</sup> Cf. por exemplo a descrição completa de um recurso no serviço ADAM, the Art, Design, Architecture & Media Information Gateway ([www.adam.ac.uk](http://www.adam.ac.uk)) que inclui, para além da descrição em texto livre, a classificação quanto ao tipo de recurso, a cobertura geográfica, a notação Dewey e informação sobre o próprio registo.

que disponibilizam essa possibilidade devem cumprir, bem como qual o papel e a função do utilizador.

Desde o final dos anos 50, quando se realizam as primeiras experiências de indexação automática, que a oposição entre este processo e o da indexação humana constitui um dos temas maiores da investigação e discussão sobre a recuperação da informação. Esta problemática reavivou-se no ambiente *Web* e, como vimos, conduziu ao aparecimento de serviços com características diversas. Os motores de pesquisa constituem, com mais incidência na sua primeira fase, o resultado da crença nas potencialidades da indexação automática. A aplicação deste processo a um universo tão extenso como a *Web* produziu resultados relativamente pouco satisfatórios, existindo experiências que demonstram uma correlação positiva entre o tipo de pergunta e a qualidade dos resultados. Assim, verifica-se que “ a melhoria na revocação e precisão que pode ser atribuída à NLP (processamento da linguagem natural) parece estar relacionada com o tipo e extensão da pergunta inicial. Perguntas mais detalhadas e mais longas respondem bem à LMI (indexação linguisticamente motivada), enquanto que pesquisas por frases breves e directas mostram poucos progressos” (CHAN: 2000) A possibilidade de expandir uma determinada questão constitui, desta forma, um dos requisitos mais importantes para assegurar um aumento de eficácia desta metodologia de recuperação.

Apesar de todos os desenvolvimentos que têm vindo a ser implementados na indexação automática, a indexação humana apresenta vantagens que os sistemas automatizados estão longe de conseguir. Em primeiro lugar, a utilização de vocabulários controlados em substituição da linguagem natural apresenta vantagens claras: “A função básica do controlo vocabular, isto é, melhor revocação através do controlo dos sinónimos e relações entre os termos, e maior precisão através do controlo dos homógrafos não têm sido completamente suplantadas pela pesquisa por palavra, mesmo com todo o poder que um sistema totalmente baseado na máquina pode produzir.” (CHAN: 2000). A indexação manual permite, na generalidade:

- Agrupar documentos sobre o mesmo assunto independentemente da língua, da terminologia ou vocabulários específicos;
- Adaptação às necessidades específicas de determinados grupos de utilizadores
- Apoiar o utilizador na definição vocabular
- Adaptação a diferentes níveis de especificidade na representação do assunto
- Recuperação consistente no tempo
- Navegação entre os assuntos, quer hierarquicamente, quer horizontalmente

As desvantagens essenciais na utilização desta metodologia e dos processos e instrumentos a ela associados prende-se essencialmente com:

- O custo, o que no ambiente *Web* é factor absolutamente crucial. Não é possível considerar a possibilidade de indexação manual para todos os recursos existentes;
- A dificuldade em manter a consistência do mesmo indexador ao longo do tempo, e muito mais, entre indexadores
- Uma certa dificuldade de adaptação à linguagem de que o utilizador se socorre para executar as pesquisas.

A dicotomia entre estes dois métodos está tendencialmente a ser resolvida, não pela exclusão de um deles, mas por vários processos de aproximação:

- a) Utilização de vocabulários controlados associados à pesquisa em texto livre, permitindo assim um alargamento ou refinamento da questão inicialmente posta pelo utilizador. (CHAN: 2000)
- b) Diversificação das possibilidades de pesquisa num mesmo serviço. O exemplo mais claro desta tendência é a progressiva expansão de directórios de assunto a serviços que inicialmente eram apenas motores de pesquisa. Mesmo um serviço recente como o Google que voltou a afirmar a superioridade da indexação automática, possui já um directório de assuntos.
- c) Utilização das potencialidades de automação para executar tarefas até agora exclusivamente manuais, como a da utilização de linguagens classificatórias para efeitos de indexação.

A prevalência histórica da indexação automática corporizada maioritariamente em serviços do tipo motores de pesquisa encontra-se, estamos em crer, relacionada com a fase do que alguns designam como a era de pré-massificação da Internet (OLIVEIRA: 2001). A eficácia da indexação automática é tanto maior quanto o corpus documental é mais específico, o que era o caso nesta fase da Internet. De facto, os problemas da sinonímia, polissemia e homografia são menos relevantes se o conjunto de documentos for mais específico e quantitativamente mais reduzido. Por outro lado, a “utopia comunicativa” que caracterizava essa época produzia uma multiplicidade de interacções entre os produtores e utilizadores de informação que se reduziam a comunidades específicas com um alto grau de entre-ajuda e de comunhão de interesses. O problema assume, de facto, uma outra escala quando se verifica o alargamento da Internet a outras comunidades, quebrando-se os laços identitários que reuniam as comunidades iniciais.

### **2.2.2. Vantagens e inconvenientes**

Como se verificou nenhum dos serviços existentes indexa o mesmo número e tipo de recursos, nem apresenta a mesma qualidade de recuperação da informação. Nesta circunstância, todos podem ser considerados relevantes dependendo do tipo de utilizador e das necessidades de pesquisa que este experimenta. Aliás, mais do que do tipo de utilizador, consideramos que a relevância de qualquer um deste tipo de serviços depende do estado de conhecimento de um dado indivíduo sobre uma questão, das suas exigências em termos de qualidade e especificidade dos conteúdos e da experiência na utilização das facilidades de pesquisa. De facto, mesmo um utilizador que aborda a *Web* para pesquisar informação sobre um “hobby” pode escolher um serviço mais generalista ou específico de acordo com o seu estado de conhecimento sobre o assunto. Esta perspectiva aproxima-se da tentativa de aplicação dos resultados do estudo realizado por Mote ao ambiente Internet levada a cabo por Wilson, quando afirma “Transferindo estas ideias para o mundo actual podemos também pôr a hipótese de que os indivíduos da categoria 1 (os especialistas num campo específico) parecem ser utilizadores mais efectivamente independentes de motores de pesquisa e sistemas de pesquisa online, enquanto que as categorias 2 e 3 (os mais generalistas) parecem continuar a necessitar de serviços de intermediários especializados” (WILSON: 2000)

Assim, é possível definir uma tipologia de situações de pesquisa e, tendo em conta as características dos vários serviços, identificar os que têm mais probabilidade de responder eficazmente: Para um utilizador com pouca experiência e sem definição exacta do assunto pretendido adequam-se melhor os directórios de assunto generalistas; para um utilizador com pouca experiência, mas com definição clara do assunto que

pretende pesquisar os directórios de assuntos ou motores de pesquisa específicos; para um utilizador com experiência de pesquisa, com clara identificação do assunto e pretendendo identificar o maior número de recursos possível os motores de pesquisa em texto completo e/ou meta motores; para um utilizador com alguma experiência de pesquisa e que pretende pesquisar assunto de carácter geral os motores de pesquisa por extracção; para um utilizador relativamente ou muito experiente, clara identificação do assunto, ao menos ao nível de uma dada temática mais ou menos abrangente, grande preocupação na validade e qualidade dos resultados, estão os *Subject Gateways* / Portais.

Independentemente da verificação da validade desta relação entre situações de pesquisa e tipo de serviço a utilizar, é necessário ter em conta que a opção por um serviço que se baseie na indexação manual ou na indexação automática tem exigências diferentes para o utilizador. No primeiro caso, as responsabilidades de garantir uma pesquisa eficaz, ou seja, a que recupera os documentos relevantes com um mínimo de ruído, é claramente transferida para este, e é isso que tem justificado o sucessivo desenvolvimento de técnicas de pesquisa nestes serviços (desde operadores booleanos, de proximidade, de restrição, etc.). Ainda neste grupo de serviços, a responsabilidade de avaliação da informação é completamente transferida para o utilizador, ou seja, nenhum motor de pesquisa garante que a informação recuperada obedece a determinados critérios de qualidade.

Ao contrário, nos serviços onde existe intervenção humana na organização da informação, sobretudo os *Subject Gateways*, o utilizador pode contar à partida com determinados padrões de qualidade, quer no que diz respeito aos conteúdos que resultam da sua pesquisa, quer no que concerne à eficácia dessa pesquisa.

### **2.2.3. Subject Gateways**

Este tipo de serviços, que cobrem assuntos específicos ou conjuntos temáticos especialmente relevantes para determinadas comunidades, maioritariamente a comunidade universitária; apostam predominantemente na intervenção humana no processo de organização da informação utilizando instrumentos standardizados e qualitativamente validados pela experiência e análise científica através de processos que visam assegurar a máxima qualidade do resultado.

Não têm ambições de identificar todos os recursos *Web* sobre a área específica em que se concentram, mas apenas aqueles que apresentam uma maior conformidade com os objectivos definidos em termos de utilizadores alvo e os critérios de qualidade definidos.

A literatura sobre *Gateways* e catalogação da Internet é vasta, apesar de esta área só ter sido considerada de forma mais global durante os anos 90. Dempsey traça uma perspectiva histórica deste tipo de serviços, os quais “emergiram em resposta ao desafio da descoberta de recursos num ambiente Internet de rápido desenvolvimento durante os anos 90 (DEMPSEY: 2000). Foi o rápido desenvolvimento da Internet, sem sistemas de classificação e indexação eficazes, que levou ao aumento das dificuldades de navegar e pesquisar a informação em rede. Dempsey refere igualmente que o termo “*Subject Garteway*” foi popularizado no âmbito do programa eLib (ver ponto 3.1.1).

A explosão da informação na Internet e a necessidade de pôr cobro ao caos potencial constituíram uma das razões pelos quais os *Gateways* se tornaram tão populares e prevalentes; como Dempsey refere existem, actualmente, “portais por todo o lado.” (DEMPSEY: 2000)

Sladen e Spence destacam as vantagens de utilização de *Subject Gateways* ao escrever: “Ao oferecerem um único ponto de acesso a recursos Internet num determinado campo, *Subject Gateways* têm uma característica essencial que os distingue de outras empresas comerciais – caracterizam-se por uma metodologia de controlo da qualidade baseada na intervenção de profissionais com origem na disciplina académica relevante.” (SLADEN e SPENCE: 2000)

Existem muitas definições do termo *Subject Gateway* na literatura, um termo que Dempsey, Gardner e Day referem “ser parte do léxico da informação em rede utilizado na investigação, ensino e área cultural.” (DEMPSEY e tal.: 1999). Koch refere que “Um *Subject Gateway* é um serviço Internet focalizado em recursos Internet distribuídos.... Que suporta a descoberta sistemática desses recursos... O serviço é baseado na descrição dos recursos. O acesso via *browsing* dos recursos utilizando uma estrutura temática é uma importante característica.” (KOCH: 2000)

Wiseman refere, por sua vez, que “*Subject Gateways* proporcionam o acesso a listas de recursos de qualidade em disciplinas específicas e oferecem uma variedade de serviços de valor acrescentado relevantes para estas disciplinas.” (WISEMAN: 1998). Dempsey, Gardner e Day sugere que os *Gateways* têm algumas ou todas das seguintes características: “serviços baseados na descrição dos recursos; alto nível de criação manual, muitas vezes, por especialistas de informação; pesquisa por palavra-chave e *browsing*; política de desenvolvimento da colecção, baseada em critérios de selecção e de qualidade; política de gestão da colecção, suportada por procedimentos de manutenção e actualização definidos.” (DEMPSEY e tal.: 1999).

Koch refere que “*Subject Gateways* com controlo de qualidade são serviços Internet aos quais se aplica um conjunto de medidas qualitativas que suportam a sistemática descoberta de recursos. Um considerável esforço manual é utilizado para assegurar a selecção de recursos de acordo com critérios de qualidade e disponibilizar descrições enriquecidas desses recursos utilizando standards baseados em metadados.” (KOCH: 2000). Estas afirmações destacam o facto de que é, em larga medida, a descrição de recursos que torna os *Subject Gateways* tão úteis. O *Desire Information Gateway Handbook* (ver ponto 3.2.1) refere que “um dos papéis chave dos *Internet Subject Gateway* é a criação de metadados descritivos sobre recursos em rede que podem ser utilizados como base para pesquisa e *browsing* do *Gateway*. Estas descrições podem também ajudar os utilizadores do *Gateway* a identificar se os recursos são exactamente o que pretendem, poupando – lhes potencialmente o tempo necessário à navegação dos recursos disponíveis em qualquer local na Internet.” (DESIRE INFORMATION GATEWAYS HANDBOOK).

No ano 2000 e no âmbito do projecto RENARDUS foi produzido um relatório sobre os aspectos estruturais dos nove *Gateways* participantes (Dutchess, Novagate, EELS, Def fagportal, Dalnet, FVL, Les Signets, RDN e SSG-FI), que por abranger uma amostra significativa, permite conhecer o universo em questão. Verificou-se que:

- a) O público-alvo comum é a comunidade académica

- b) A abrangência temática da colecção é vasta
- c) Em termos de cobertura geográfica, os recursos têm origem sobretudo na Europa e nos E.U.A.
- d) Todos os *Gateways* têm mais ou menos o mesmo tipo de critérios de selecção, a saber: conteúdo, processo, forma e cobertura temática
- e) A maioria dos recursos seleccionados são sítios *Web*, de acesso livre e/ou de domínio público
- f) A granularidade é, mais ou menos, ao nível da colecção / serviço
- g) As linguagens de metadados baseiam-se ou nas Regras de Catalogação ROADS ou em Dublin Core. Só dois *Gateways* usam linguagens desenvolvidas especificamente para o próprio serviço. Estas linguagens são sempre aplicadas por especialistas. O mapeamento entre as várias linguagens demonstrou que todos descrevem os seguintes atributos: **título, autor, descrição e identificação (URL)**
- h) A maioria dos *Gateways* usa regras de descrição baseadas em regras formais de catalogação
- i) A maioria dos *Gateways* não utiliza ficheiros de autoridade
- j) Todos os *Gateways* proporcionam a funcionalidade de *browsing*, na maioria dos casos, baseada num esquema de classificação
- k) Quase todos os *Gateways* disponibilizam funcionalidades de pesquisa por palavra, simples e avançada
- l) Alguns *Gateways* permitem a combinação de *browsing* e pesquisa por palavra
- m) Vários mecanismos de controlo de qualidade são utilizados nomeadamente verificação das ligações e controlo editorial (BECKER: 2000)

Recentemente Howarth situando os *Subject Gateways* como produtores de formatos de metadados enriquecidos, e definidos de acordo com standards internacionais, sugere os seguintes atributos da informação a ser incluídos em qualquer estrutura de metadados definidos por qualquer *Gateway* (excluem-se os metadados de administração):

- a) Metadados descritivos: descrevem as propriedades físicas e intelectuais de um objecto digital:
  - Título (incluí títulos alternativos, títulos paralelos, complementos de título, títulos abreviados, etc.)
  - Criador (autor, compositor, artista, etc.)
  - Data
  - Editor
  - Identificador único e ligação dinâmica (URI, URL, etc.)
  - Sumário, nota descritiva, revisão, etc.
  - Nível de audiência
  - Média físico, formato, etc.
- b) Metadados analíticos: informação que analisa e acrescenta valor ao acesso ao conteúdo do recurso:
  - Cabeçalhos de assunto
  - Tesouros
  - Palavras-chave
  - Resumo
  - Notações a partir de sistemas de classificação
  - Outros elementos de importância local, como por exemplo, afiliação departamental, ligação a outros recursos, etc.

- c) Metadados para gestão dos direitos de acesso: informação relativamente a restrições (legais, financeiras, etc.) ao acesso e utilização dos objectos digitais:
  - Restrições à utilização
  - Declarações de permissão de acesso
  - Taxas de subscrição, licenciamento ou utilização
  - Informação sobre direitos de autor
  - Escalas de qualidade (*quality ratings*)
- d) Metadados técnicos: hardware e software utilizado na conversão do objecto no formato digital, no seu armazenamento, visualização, etc.:
  - Especificações do equipamento de digitalização
  - Posições da camera
  - Condições de captura da imagem
  - Parâmetros de codificação
  - Reconhecimento da voz
  - Especificações de scanner óptico
  - Tipo de ficheiro e requisitos do software de conversão (HOWART:2003)

Os *Gateways* aplicam tendencialmente, como teremos oportunidade de verificar com a análise do caso dos *Gateways* da rede RDN, um modelo de metadados descritivos que se apresenta como uma tentativa de adaptação do modelo tradicional da descrição ao universo dos recursos em rede. A aplicação deste modelo descritivo visa a concretização dos objectivos universais que qualquer utilizador espera de um sistema de organização da informação, na medida em que, desde logo, a comunidade a que se destinam espera a concretização dos referidos objectivos. (cf. Ponto 2.3)

Em síntese, estes serviços apresentam as seguintes características básicas:

### **1. Selecção de Recursos**

Um intermediário humano adiciona valor seleccionando recursos Internet apropriados, normalmente de acordo com critérios de selecção rígidos. Os recursos são normalmente seleccionados pela sua qualidade, autoridade, actualidade e relevância do assunto. Outros critérios de selecção também podem ser aplicados tais como língua ou cobertura geográfica.

### **2. Gestão da colecção**

Realiza-se uma manutenção regular da colecção, incluindo a remoção de recursos que já não são apropriados, foram suprimidos ou contém erros na entrada dos dados. Verificação regular das ligações pode também ser levada a cabo. Estes procedimentos podem ser automatizados ou levados a cabo manualmente através de intervenção humana. Deve existir uma política que detalhe estes procedimentos.

### **3. Descrição dos Recursos**

Os recursos seleccionados são anotados por um intermediário humano através de uma descrição completa do recurso. As descrições serão produzidas de acordo com um esquema de metadados estruturado e pré-definido. Os metadados são estruturados em campos separados. Isto permite que o recurso seja facilmente identificado e localizado. É também facilitador de pesquisas estruturadas. As descrições podem conter informação sobre o conteúdo, autor, editor ou data de publicação do recurso.

#### **4. Classificação do Assunto**

Um intermediário humano utiliza um esquema de classificação para indexar todos os recursos. Isto facilita o *browsing* por assunto. Os *Subject Gateways* são também caracterizados pela aplicação de standards que permitem a interoperabilidade com outros serviços, bem como por várias funcionalidades de valor acrescentado.” (CAMPBELL:2001)

A importância e eficácia destes serviços face a outros para a comunidade a que se destinam tem vindo a ser reforçada e já em 1998 um estudo comparativo sobre *Subject Gateways* recomendava a sua continuação e reforço através de uma estratégia de coordenação da cobertura temática, da criação de um único ponto de acesso a todos os recursos, da racionalização dos custos e procura de financiamentos alternativos e eventual comercialização de alguns serviços. (HAYNES: 1998)

Num estudo posterior no contexto do projecto RENARDUS que pretendia verificar a forma como os utilizadores usavam este tipo de serviços, conclui-se que estes referem como aspectos positivos dos *Subject Gateways*:

- a) A qualidade dos recursos
- b) As opções de pesquisa e *browsing*, sendo a primeira utilizada para encontrar um número mais alargado de recursos e a segunda para pesquisas mais específicas. Dentro desta última opção privilegiam a pesquisa simples e os pontos de acesso palavras-chave, autor, título e descrição em texto livre
- c) Qualidade das descrições dos recursos
- d) Variedade dos recursos (BECKER: 2000a).

Os *Subject Gateways* têm assim revelado junto da comunidade académica e de investigadores um importante papel no que diz respeito ao acesso à *Web*. Dochartaigh explica porque razão estes serviços são tão úteis para este tipo de público – alvo: “ A investigação envolve a exploração de uma área relativamente vasta do conhecimento, uma exploração do que foi escrito e do que está a ser escrito nessa área. O melhor ponto de partida nessa exploração na *Web* são os guias de assunto.” (DOCHARTAIGH:2002, p. 57)

### **2.3. Sistemas de organização da informação**

Organizar a informação, como condição central para assegurar a recuperação eficaz, constituiu uma tarefa socialmente reconhecida desde, praticamente, o início da cultura escrita e ganhou, como já tivemos oportunidade de demonstrar, uma nova relevância com a explosão do universo informativo do final do século XX. Mas, o que significa exactamente este conceito de organizar a informação e que objectivos se pretendem atingir, sobretudo do ponto de vista dos utilizadores?.

Organizar informação apresenta importantes diferenças relativamente a outros processos organizativos. Assim, a aplicação de procedimentos normalmente utilizados para desenhar bases de dados para organizar entidades como alunos, stocks, etc. não podem ser aplicadas da mesma forma. A primeira grande diferença reside na necessidade de organizar ao mesmo tempo e de forma relacionada duas entidades distintas, que são as obras (entidades intelectuais com vários invólucros possíveis) e os documentos que as corporizam. (SVENONIUS: 2001, p.10).

A forma mais comum de organizar é classificar. A classificação reúne entidades que participam de uma ou mais características comuns, definidas de acordo com vários critérios. O critério essencial em sistemas de informação é o atributo “corporiza a mesma informação” (SVENONIUS: 2001, p.10). Assim, organizar significa em primeira instância reunir exactamente a mesma informação. Para além disso também se pretende reunir informação semelhante, sendo que este conceito é de difícil tradução num código de regras aplicável a todo o universo informacional.

O objectivo essencial de um sistema de organização da informação é, assim, reunir informação semelhante e distinguir as dissemelhanças. O desenho de um sistema para atingir este objectivo tem que fazer face a vários problemas, entre os quais se destacam: a infinita variedade do universo da informação e a definição de uma linguagem para aceder à informação. A utilização da linguagem natural para este desiderato levanta problemas de recuperação ligados, sobretudo, às suas características de polissemia e sinonímia, provocando por isso ambiguidades e redundâncias que dificultam a recuperação. Assim, a tarefa mais importante na construção de um sistema de organização da informação é a de criar uma linguagem da descrição “que imponha sistemática e método à linguagem natural e que, ao mesmo tempo, permita aos utilizadores encontrar o que pretendem pelos nomes que conhecem.” (SVENONIUS: 2001, p.14)

### **2.3.1. Objectivos de um sistema de informação bibliográfica**

A definição de objectivos concretos para um sistema de organização da informação teve a sua primeira grande concretização com Cutter em 1876. Em 1960 Lubetzky reviu os objectivos tradicionais estabelecendo a distinção entre Obra e livro e afirmando a primazia do conteúdo da informação como atributo de classificação. Em 1997, no estudo da IFLA intitulado “*Functional requirements for bibliographic records*” faz-se uma nova revisão dos objectivos estabelecendo os seguintes: encontrar entidades que correspondam ao um determinado critério de pesquisa do utilizador, identificar uma entidade, seleccionar uma entidade apropriada às necessidades do utilizador, adquirir ou obter acesso à entidade descrita (Functional requirements for bibliographic records: 1997, pp. 8-9).

Elaine Svenonius analisa estes objectivos argumentando que a inclusão de dois objectivos tradicionais (encontrar e reunir) no mesmo objectivo inicial diminui a importância do objectivo reunião, um dos mais importantes para qualquer utilizador. Por outro lado, no modelo IFLA está ausente, pelo menos de uma forma explícita, o objectivo de navegação no universo bibliográfico, o que tendo em conta, por um lado, a diversidade de suportes hoje existentes, e por outro a facilidade tecnológica no estabelecimento de ligações entre entidades, se torna cada vez mais premente. Assim, a autora que vimos seguindo redefine os objectivos, que podemos designar por universais, de qualquer sistema de informação bibliográfica da seguinte forma:

1. Localizar entidades como resultado de uma pesquisa que utilize atributos ou relações entre as entidades, subdividido em:
  - 1.1. Localizar uma entidade singular
  - 1.2. Localizar conjuntos de entidades que representem:
    - 1.2.1. Todos os documentos pertencendo à mesma **Obra**

- 1.2.2. Todos os documentos pertencendo à mesma edição
- 1.2.3. Todos os documentos de um dado autor
- 1.2.4. Todos os documentos sobre um determinado assunto
- 1.2.5. Todos os documentos definidos por outros critérios
2. Identificar uma entidade (quer dizer, confirmar que a entidade descrita num registo corresponde à entidade que se pensava existir ou distinguir entidades com características similares)
3. Seleccionar uma entidade apropriada às necessidades do utilizador (quer dizer, escolher uma entidade que corresponda aos requisitos do utilizador no que diz respeito a conteúdo, formato físico, etc., ou rejeitar uma entidade desapropriada para responder às necessidades do utilizador)
4. Adquirir ou obter acesso à entidade descrita (quer dizer, adquirir uma entidade através de compra, empréstimo ou aceder a uma entidade electronicamente através de uma conexão online)
5. Navegar no universo bibliográfico (quer dizer, encontrar obras relacionadas com outras por generalização, associação, e agregação; encontrar atributos relacionados por equivalência, associação e hierarquia). (SVENONIUS: 2001, p.20)

Como afirma Svenonius, apoiada em estudos sobre necessidades do utilizador, os cinco objectivos definidos podem ser considerados como uma afirmação daquilo que os utilizadores têm o direito de esperar de sistemas de organização da informação” (SVENONIUS: 2001, p.29)

### **2.3.2. Descrição dos recursos de informação**

Nos sistemas de organização assim definidos, a descrição dos recursos de informação é a base de todo o processo de organização de acordo com o princípio de que para organizar é preciso descrever. O objecto da descrição são as entidades de informação consideradas na sua dupla vertente de conteúdo e invólucro físico.

Ao descrever-se um objecto de informação pretende-se comunicar, através de uma linguagem específica, desenhada e aplicada de acordo com um conjunto particular de regras, determinados atributos dessa entidade. Esta linguagem é artificial, quer dizer expurgada das anomalias da linguagem natural, nomeadamente os problemas levantados pelas características de sinonímia, polissemia e homografia, pretendendo proporcionar um acesso sistemático à informação, reduzindo a aleatoriedade da ausência de organização.

A linguagem da descrição produz entidades informativas que podem ser designadas por descrições e definidas como um conjunto de afirmações das propriedades de uma coisa ou das suas relações com outras coisas que servem para a identificar. Como qualquer linguagem, também as deste tipo se compõem de: vocabulário, consistindo nas expressões usadas para nomear os valores das variáveis e entidades a descrever, atributos dessas entidades e relações que estabelecem entre si; semântica, ou seja, as relações que se estabelecem entre os nomes; sintaxe, quer dizer, um conjunto de regras que estabelece a ordenação das relações entre os elementos que compõem expressões complexas e, finalmente, uma pragmática, ou seja, a definição das condições de aplicação da linguagem.

O vocabulário pode ser definido como uma colecção de termos ou códigos disponíveis para utilização (SVENONIUS: 2001, p. 55) e são designados por vários nomes: nas linguagens dos assuntos como palavras-chave, termos de indexação ou notações; nas linguagens da **Obra** e do **Documento** são referidos elementos de dados ou metadados.

As regras que governam o vocabulário de uma linguagem bibliográfica estabelecem critérios para a sua utilização, por exemplo a fonte a partir da qual esse vocabulário deve ser extraído e a sua forma. O objectivo das regras de forma é assegurar consistência na descrição, especificando como seleccionar entre formas alternativas de um nome, a forma linguística desse nome, o seu estilo de abreviação e capitalização e a língua e a escrita a ser utilizada.

O vocabulário é constituído por dois tipos de termos: termos derivados e termos atribuídos, sendo os primeiros derivados da fonte prescrita da informação e os segundos atribuídos a partir de uma dada normalização.

As regras que definem a semântica nas linguagens bibliográficas estabelecem três tipos de estruturas: semântica relacional, semântica referencial e semântica categorial. A implementação destas estruturas implica normalizar a linguagem natural através da introdução de vocabulário controlado ou de controlo de autoridade.

A semântica relacional trata do sentido das relações entre os termos, nomeadamente as relações entre sinónimos, as relações hierárquicas e as conexões entre termos relacionados. A semântica referencial lida com um conjunto de técnicas utilizadas para limitar o sentido dos termos, utilizando normalmente qualificadores que permitem distinguir entre entidades semelhantes. A semântica categorial estabelece as regras de partição do vocabulário em facetas ou categorias gramaticais.

Os dois grandes tipos de linguagens destinam-se a descrever as duas grandes entidades do universo informativo: **Obras** e **Documentos**. Estas linguagens podem ainda ser subdivididas de acordo com os atributos seleccionados para a descrição. Assim, a linguagem da **Obra** pode ser subdividida em Linguagem do autor, Linguagem do título, Linguagem da edição e Linguagem do assunto, esta última podendo ainda subdividir-se em Linguagem de classificação e Linguagem de indexação. Para a entidade **Documento** é possível distinguir os seguintes tipos de linguagens: Linguagem da produção, Linguagem da localização e Linguagem do invólucro físico. Em alguns casos, sobretudo autores e assuntos, estas entidades são também consideradas atributos de outras entidades.

As linguagens da **Obra** descrevem entidades de informação, os seus atributos intelectuais e as relações que se estabelecem entre elas. Entre os atributos da **Obra** contam-se os **assuntos**, os **títulos**, as **pessoas físicas** ou **colectivas** responsáveis pela sua criação, a **data** e o **contexto do público-alvo**, o simbolismo que usam para comunicar informação (tais como lexical, musical ou gráfico) e os sentidos requeridos para receber essa informação. Do ponto de vista da identificação e estruturação da informação os atributos mais fundamentais são os assuntos e os atributos nominais, ou seja, os que servem para nomear obras e seus criadores.

Nas linguagens tradicionais da descrição utilizam-se dois tipos de expressões para nomear atributos: as que derivam directamente da informação e as que são atribuídas por quem descreve. As primeiras usam vocabulário não controlado, as segundas vocabulário controlado. As primeiras asseguram que as descrições representam efectivamente os documentos que pretendem descrever; as segundas são construções de linguagem artificial; o seu objectivo é mapear o vocabulário dos utilizadores com um vocabulário standardizado e reunir informação semelhante. A utilização de vocabulário controlado é condição essencial para assegurar um dos objectivos essenciais dos sistemas de organização de informação: reunir informação semelhante, ou seja, localizar conjuntos de entidades que partilhem a mesma instanciação de um dado ponto de acesso: autor, assunto, título, etc.

A linguagem do **Documento** descreve e permite acesso ao invólucro físico que corporiza a informação e destina-se a apoiar a concretização dos seguintes objectivos: encontrar um documento, identificar um documento, seleccionar um documento, adquirir um documento e navegar entre documentos.

Os atributos descritos por este tipo de linguagens inscrevem-se em três categorias: atributos físicos (**media** e outros atributos físicos), atributos de publicação (**editor, local e data**) e atributos de acesso (**localização, condições de acesso**).

A diversidade do universo dos objectos de informação, dos públicos-alvo, bem como os objectivos que se pretendem atingir podem tornar estas regras muito complexas, aumentando os custos de aplicação e a viabilidade da sua extensão a todo o universo informativo. Assim, a discussão sobre o nível da descrição a utilizar, introduzindo muitas vezes o conceito de nível mínimo, é já tradicional no ambiente das bibliotecas, tendo vindo recentemente a tomar um maior fôlego a propósito dos recursos *Web*.

Em favor de uma descrição minimalista jogam argumentos como o custo e alguns estudos de utilizadores que apontam para a incapacidade de muitos destes em explorar todas as possibilidades conferidas por estes sistemas complexos, considerando-se que este tipo de sistemas não é necessário para cumprir todos os objectivos pretendidos. De acordo com a avaliação de Svenonius, estes estudos foram feitos sobre amostras muito limitadas pelo que as suas conclusões são muito relativas. (SVENONIUS: 2001, p. 28).

Em favor de uma descrição completa militam igualmente argumentos baseados no utilizador que sustentam que para muitos tipos de utilizadores, nomeadamente os da comunidade escolar e científica, uma descrição completa assegura o objectivo reunião de conjuntos, essencial para o desenvolvimento da investigação. Por outro lado, estudos sobre os comportamentos de pesquisa dos utilizadores apoiam também uma descrição mais completa e sistemas mais performantes, já que demonstram que estes têm dificuldade em encontrar termos de pesquisa apropriados, em expandir uma pesquisa quando muito pouco é recuperado e em limitar uma pesquisa quando os resultados são quantitativamente extensos. (SVENONIUS: 2001, pp.28-29).

## 2.4. Metadados

Os recursos *Web* deram origem a uma proposta de linguagens da descrição que se apresentaram, pelo menos inicialmente, como alternativa às linguagens já existentes.

Numa primeira aproximação ao conceito de metadados é usual encontrar a expressão de que metadados são dados sobre dados ou informação sobre informação, o que não é particularmente esclarecedor. O prefixo meta, de origem grega, pode indicar o conceito de mudança como no termo metamorfose ou pode significar para além de / depois de como em metafísica. É mais nesta última acepção que ele é utilizado no contexto dos recursos de informação presentes na *Web*, como produção de informação descritiva sobre os recursos informativos disponíveis. Trata-se então de um conjunto de informação, ou seja, dados, que visam descrever as características dos recursos com vista a facilitar a sua recuperação. Cada um dos dados sobre o recurso é independente de outros e podem ser classificados em dois tipos: dados descritivos propriamente ditos e ligações.

Tim Berners-Lee, acrescenta a este conceito generalista uma característica importante que deve possuir a informação descritiva sobre o recurso *Web*: é informação para ser processada e entendida por máquinas. Como ele próprio afirma: “A frase “entendível pela máquina” é chave. Estamos a falar de informação que aplicações de software podem usar por forma a tornar a nossa vida mais fácil, assegurar que obedecemos aos nossos princípios, à lei, verificar que podemos confiar no que estamos a fazer e tornar o trabalho de todos mais leve e fácil.” (BERNERS-LEE:1997). Esta definição permite esclarecer os objectivos da criação de metadados: recuperação da informação que circula em ambiente *Web*, gestão de aspectos legais como os direitos de autor, assegurar um certo nível de qualidade e fiabilidade da informação (o Projecto PICS, *Platform for Internet Content Selection*, constitui um exemplo para um aspecto específico desta última funcionalidade possível dos metadados). Assim, podemos afirmar, de acordo com Berners-Lee, que: “Metadados são dados. Os metadados podem referir-se a um recurso que tem um URI. Metadados podem ser considerados como um conjunto de afirmações, cada afirmação referindo-se a um recurso. Afirmações que nomeiam uma relação entre dois recursos são conhecidas como ligações. Os tipos de afirmações devem ser considerados objectos de primeira classe no sentido de que devem poder ser definidas em objectivos endereçáveis e referidas pelo endereço desse recurso. O desenvolvimento de novas afirmações e relações deve ser feito de forma consistente para que este tipo de afirmações possa ser tratado na generalidade por pessoas e software.” (BERNERS-LEE: 1997).

Para além do objectivo primordial de contribuir para a recuperação dos recursos *Web*, a noção de metadados, enquanto meta-informação, aponta claramente para a descrição propriamente dita, ou seja, o conjunto de informação produzida com vista a facilitar essa mesma recuperação. Metadados devem, então, constituir-se como uma dada semântica de descrição dos recursos informativos. A sintaxe desta semântica não é, na grande maioria dos casos, constituinte da estrutura de metadados; visto que esta necessidade está mais relacionada com a operacionalização da recuperação; a sintaxe é dada pelo formato de representação dos metadados (HTML, XML, RDF por exemplo.)

O nível de descrição previsto em cada linguagem de metadados é diverso e vai desde metadados pouco estruturados que não utilizam qualquer tipo de controlo vocabular até

metadados altamente estruturados, utilizando diversos processos de controlo terminológicos para vários atributos, por exemplo, **autores** e **assuntos**. Dempsey e Heery propõem uma classificação dos metadados utilizando exactamente o critério da riqueza estrutural e semântica e sua complexidade, agrupando as linguagens de metadados em três bandas que evoluem do menos complexo para o mais complexo: a Banda 1 é composta pelos formatos simples criados em sistemas proprietários e pressupõem a indexação em texto livre, ou seja, os metadados gerados predominantemente de forma automática e utilizados no ambiente dos motores de pesquisa (também conhecidos por pseudo-metadados); a Banda 2 composta por formatos estruturados que utilizam standards e uma estrutura de campos de informação; finalmente a Banda 3 inclui os formatos mais ricos que utilizam standards internacionais e uma elaborada estrutura de etiquetas para incluir a informação sobre os atributos. (DEMPSEY: 1997). A geração destas descrições pode ser realizada de forma completamente automática, completamente manual ou através do que poderíamos designar por geração assistida de metadados, quer dizer, aquela que utiliza alguns instrumentos que extraem automaticamente determinados atributos da informação, a serem complementados e/ ou validados manualmente. Os criadores destas descrições podem oscilar entre os produtores não especializados da informação até aos especialistas em gestão da informação.

Tendo em conta o objectivo de facilitar a recuperação da informação e a impossibilidade, dadas as características da Internet, de assegurar que este tipo de informação possa ser sistematicamente produzido sobre todos os recursos por uma ou várias organizações mundiais, os metadados pressupõem a sua presença na própria informação, ou seja destinam-se a ser "embebedos" nas páginas *Web*, e a ser inscritos nessas páginas pelo próprio produtor da informação. Esta forma de operacionalizar a descrição das características dos recursos é hoje complementada com a prática de construção de verdadeiros catálogos de recursos, sobretudo para os casos em que a complexidade semântica das linguagens de metadados utilizadas obriga a uma gestão dos conteúdos por especialistas de gestão de informação no contexto, por exemplo, de serviços de pesquisa na *Web*.

Quer para a descrição de recursos *Web* quer para as implementações de projectos de bibliotecas digitais, existe uma grande diversidade de estruturas de metadados empregues. Sem cair num grande exagero podemos afirmar que existem tantas estruturas de metadados quanto projectos de bibliotecas digitais, tantas propostas quantos os campos científicos abordados pelos documentos produzidos para a *Web*. Milstead e Feldman chegam mesmo a adjectivar esta realidade como o dilema dos metadados. (MILSTEAD: 1999). Esta diversidade impede em grande medida a interoperabilidade, condição essencial para assegurar a eficácia da recuperação, sobretudo num ambiente distribuído como o da *Web* e onde os documentos têm as mais variadas origens.

A miríade de propostas de metadados actualmente existente explica-se, por um lado, pela recente utilização deste conceito e por outro, pela dificuldade em harmonizar os interesses e as necessidades de diversas comunidades de produtores e utilizadores de informação. No entanto, é um dado adquirido que esta situação não deve manter-se por muito mais tempo: surgem, cada vez com mais frequência, propostas de estruturas de metadados que possam ir ao encontro das necessidades básicas de todos e começa a trilhar-se o caminho para a normalização, recentemente confirmado com a

transformação de Dublin Core num standard ISO. Em determinados domínios, por outro lado, standards mais dominantes começam a estender-se a toda a área, como é o exemplo de ONIX para as publicações comerciais e LOM para a educação; em muitas destas áreas desenvolve-se também a relação e interoperabilidade entre standards específicos e Dublin Core; verifica-se, por último, uma mudança de enfoque de esquemas mais casuísticos para abordagens mais abrangentes de modelação dos dados como é o caso do modelo IFLA FRBR<sup>7</sup> e do modelo ABC Ontology<sup>8</sup>. (CAUSTON: 2002) Este caminho está, de alguma forma, facilitado pela existência de componentes semânticos iguais em todas as estruturas actuais: é o caso do autor(es) (quer se designem por agentes, criadores ou autores), título, assunto(s), localização "virtual". Da mesma forma, verifica-se a tendência para a aplicação de dois princípios básicos na definição destas estruturas: modularidade, ou seja, a possibilidade de combinar elementos com origens diferentes integrando-os nos requisitos funcionais de um dado perfil de descrição; diversos níveis de granularidade possíveis, ou seja, a descrição pode fazer-se ao nível do conjunto (um recurso como tal enquanto conjunto de páginas ou uma colecção de recursos) ou de cada uma das partes.

A partir da análise empreendida, as linguagens de metadados podem ser classificadas de acordo com os seguintes critérios:

- a) Considerando a fonte e o momento de criação, podemos distinguir, por um lado, metadados internos, gerados para um DLO no momento da sua criação e/ou digitalização, e, por outro, metadados externos, relacionados com o **Documento** como objecto, mas criados à posteriori por alguém distinto do agente original que cria a informação
- b) Segundo o método concreto de criação, podemos distinguir metadados gerados automaticamente e metadados criados manualmente.
- c) Se tivermos em conta o nível de especialização dos responsáveis no processo de atribuição dos metadados, podemos identificar também dois tipos: aqueles que são criados por pessoas que não são especialistas em informação e aqueles que são criados por especialistas em informação
- d) Relacionado com este último aspecto, podemos utilizar outro factor de classificação: o controlo semântico dos dados. Assim, podemos distinguir: metadados controlados que utilizam um vocabulário standard e metadados não controlados
- e) Segundo a forma de armazenamento podem distinguir-se metadados embebidos no próprio documento e metadados separados do documento
- f) Segundo o nível de estruturação, existe uma gradação da metainformação que vai desde metadados altamente estruturados a metadados não estruturados
- g) Segundo o nível da descrição, podem distinguir-se metadados de colecção, que realizam um registo ao nível da colecção, e metadados individuais, que tomam como referência uma unidade digital mínima (MÉNDEZ RODRÍGUEZ: 2002, pp. 69-71)

A origem do termo *metadata* parece estar em Jack E. Myers e na *Metadata Company* que registou este termo em 1986. A sua chegada ao ambiente *Web* dá-se por volta do início dos anos 90, quando se tornou evidente a necessidade de descrever os recursos de

---

<sup>7</sup> FRBR (Functional requirements for bibliographic records). Disponível em <http://www.ifla.org/VII/s13/frbr/frbr.html>

<sup>8</sup> Cf. Informação disponível sobre este modelo em <http://jodi.ecs.soton.ac.uk/articles/vo2/io2/lagoze/>.

informação existentes Esta evidência nasceu, em grande medida, das dificuldades reveladas pelos motores de pesquisa em assegurar uma recuperação eficaz da informação. Assim, como as bibliotecas já tinham, aliás, descoberto há mais de um século:

"Encontrar documentos requer descrever documentos. Encontrar documentos requer descrever pessoas." (MILLER: 2000); e poderíamos acrescentar: Encontrar documentos implica descrever o seu conteúdo; Encontrar documentos implica identificar a sua localização, etc.

Na descrição dos objectos digitais reside uma das questões de maior problemática no domínio da organização da informação neste ambiente. Por um lado, apresentam-se as linguagens que apostam na simplicidade descritiva como Dublin Core, no lado oposto a utilização directa da tradição descritiva corporizada, entre outras, pelas regras da descrição para recursos electrónicos (ISBD (ER))<sup>9</sup> que aplicam os princípios descritivos tradicionais.

A tendência para descrições de características minimalistas tem vindo a ser, de alguma forma, contrariada dentro do próprio campo dos metadados não só pela criação de um cada vez maior número de perfis aplicativos mais complexos para comunidades específicas, como pela possibilidade de desenvolvimento de linguagens simplificadas, como é o caso do desenvolvimento de qualificadores em Dublin Core. O principal argumento a favor desta tendência é o de que desta forma é possível aproximar melhor o conteúdo descritivo das necessidades de utilizadores específicos. Argumentando de forma contrária Carl Lagoze aponta com clareza a inadequação desta tendência tomando como base as características do modelo Dublin Core (DC) e os princípios que orientaram a sua criação. Assim, quer do ponto de vista do modelo, quer do vocabulário, quer ainda do processo de criação de qualificadores, o autor considera que o DC não é apropriado para constituir a base para descrições mais complexas. Lagoze demonstra, aliás, com um exemplo, os problemas do desenvolvimento dos qualificadores, realçando que a essencial natureza plana do modelo DC não é apropriada para descrições que implicam o estabelecimento de relações estruturantes entre os elementos. Assim, Lagoze conclui que "tentando esconder a complexidade num conjunto de metadados simples como o DCMES (Dublin Core *Metadata element Set*) conduz a compromissos inaceitáveis. Os metadados resultantes são inadequados para objectivos de simples descoberta – os princípios de *dum-dow* não funcionam – e insuficientes para descrições complexas, que requerem definições claras das entidades e suas propriedades." (LAGOZE: 2001). Enquanto isto, outros argumentam que o DC não está completo, sobretudo no que diz respeito aos qualificadores. Martin Dillon afirma claramente que o DC está "obviamente incompleto, sem qualificadores para os elementos Criador, Contribuinte e Editor." (DILLON: 2000)

Descrição dos recursos, ou catalogação, constitui uma prática biblioteconómica tradicional. Mas como Hsieh-Yee refere, "a natureza e variedade de recursos na Internet levantam a necessidade de organização da informação na era digital... a questão está em saber se a catalogação é a resposta." (HSIEH-YEE: 2000). Medeiros não tem dúvida que a catalogação bibliográfica não é a resposta (MEDEIROS: 1999). Dempsey, por seu lado, sugere que "o que é necessário é uma descrição que seja simples de criar mas desenvolvida o suficiente para proporcionar uma recuperação efectiva e uma avaliação.

---

<sup>9</sup> International standard bibliographic description: electronic resources. Disponível em <http://www.igla.org/VII/s13/pubs/isbd.htm>

Isto implica uma descrição que esteja entre a ligeireza dos motores de pesquisa e a profundidade dos catálogos das bibliotecas especializadas. “ (DEMPSEY: 2000).

Do lado dos apologistas da aplicação do modelo tradicional, Gorman apresenta uma perspectiva mais radical começando por reafirmar a necessidade de seleccionar, em primeiro lugar, que tipo de recursos têm interesse e validade para serem objecto de um tratamento profissional, identificando em seguida o que define como sendo as três perspectivas possíveis para encarar a descrição:

- a) “Identificá-los e catalogá-los de acordo com standards utilizados para outros materiais
- b) Identificá-los e usar uma parte do formato MARC, chamar-lhe “metadados”, se isso nos fizer sentir melhor, e preenchê-lo com conteúdos de acordo com standards bibliográficos
- c) Identificá-los, utilizar uma parte do formato MARC e deixar qualquer um o preencha com qualquer conteúdo “

Gorman conclui, então, que os metadados “tal como são concebidos actualmente evoluirão no sentido da standardização dos elementos e do conteúdo e tornar-se-ão pouco diferentes da catalogação num curto espaço de tempo.” (GORMAN: 2002, p.181)

Entre uma e outra perspectiva, têm vindo a desenvolver-se abordagens que propõem o que, de alguma forma, poderíamos apelidar de “terceira via”, traduzindo-se na tentativa de aplicar o modelo tradicional da descrição a estes novos recursos e ambientes. Entre estes encontram-se a obra de Nancy Olson “Cataloguing Internet Resources: a manual and practical guide” (1997)<sup>10</sup>, publicado no âmbito do projecto “Internet Cataloguing” da OCLC que começou em 1991, e o pacote aplicacional e regulamentador produzido no âmbito do projecto ROADS (ver ponto 3.2.2).

Quase na mesma linha, Medeiros argumenta a favor de uma abordagem colaborativa à questão da catalogação da Internet; uma conjugação de novas e velhas técnicas: “A complementaridade das práticas tradicionais (OPACS / MARC / AACR) com standards emergentes (METADATA) é, na minha opinião, um caminho razoável a prosseguir. Actualmente, o catálogo pode estar desactualizado, mas as suas vantagens não.” (MEDEIROS: 1999).

Méndez Rodríguez argumenta na mesma linha de colaboração, indicando a necessidade de adaptar os instrumentos de organização tradicionais ao novo ambiente: “defendemos a postura da criação de metadados, qualquer que seja a forma em que estes sejam atribuídos e o formato ou esquema que se adopte tomando aqueles aspectos da catalogação tradicional que podem contribuir para uma melhor gestão da metainformação, e, por conseguinte, da recuperação da informação. (...) É necessário um processo de adaptação intelectual para aplicar os modelos abstractos, as regras, as práticas, os códigos e os procedimentos da biblioteconomia e documentação às características das publicações electrónicas. Os metadados cobrem o vazio entre os extremos do tratamento e recuperação da informação na Internet: o da catalogação minuciosa, completa e tradicional dos recursos electrónicos e o seu acesso através de motores de pesquisa. “ (MÉNDEZ RODRÍGUEZ: 2002, pp. 87-88)

---

<sup>10</sup> Disponível em <http://www.oclc.org/oclc/man/9256cat/toc.htm>

O confronto entre vários modelos com todas as “terceiras vias” possíveis pode tornar-se numa discussão tecnicista de oposição entre tradição e modernidade que esquece, a nosso ver, a questão essencial que é a de saber que expectativas têm os utilizadores perante sistemas que pretendem, qualquer que seja a forma, organizar estes novos recursos de informação. Raramente esta questão é desenvolvidamente tratada. Méndez Rodríguez é dos poucos autores que ensaia uma definição dessas expectativas definindo-as desta forma:

- a) “O utilizador necessita de conhecer a existência de recursos, quer dizer, recuperar ou descobrir
- b) O utilizador que utiliza a rede para investigação precisa de uma informação mais ampla sobre um recurso, para além do que pode ser uma mera descrição, a saber: necessita de conhecer a origem, história ou responsabilidade intelectual da fonte, a sua integridade e autenticidade, as relações com outros recursos, as características específicas do domínio ao qual se aplica a informação, etc. Precisa também de saber se um recurso se ajusta a diferentes níveis de utilização; se pode extrair o conteúdo com alguma ferramenta ou aplicação específica; conhecer o tipo de publicação.
- c) Como futuro produtor de informação num processo de retroalimentação constante na produção do conhecimento, o utilizador necessita também de ter informação sobre os direitos de propriedade intelectual vinculados a essa informação, e que nível de utilização ou reprodução pode fazer dela
- d) Para além disso, o utilizador como consumidor de informação necessita saber sob que condições pode aceder a um determinado objecto de informação, por exemplo se tem que pagar pelo acesso. O utilizador cliente tem que conhecer também os interfaces técnicos, os protocolos de acesso, os tipos ou formatos de pesquisa, etc.
- e) Por último, também precisa de uma valorização sobre os conteúdos”. (MÉNDEZ RODRÍGUEZ:2002, p. 49)

Esta definição peca por falta de clareza, para além de ser excessivamente enumerativa, portanto, casuística; donde sem a necessária abstracção que permita adaptá-la a circunstâncias diferentes. O que significa exactamente descobrir recursos? Que atributos da informação espera o utilizador poder usar para chegar ao recurso? Descobrir implica ou não seleccionar? Se sim, que informação necessita o utilizador de ter disponível para seleccionar um recurso como o mais apropriado às suas necessidades? Que tipo de relações entre os recursos podem ser tipificadas? Elas devem estabelecer-se entre os recursos em si ou entre a meta-informação sobre cada recurso? Que tipo de informação deve estar disponível para permitir ao utilizador aceder directamente à entidade descrita?

A definição clara e o mais objectiva possível das expectativas dos utilizadores de um sistema de informação de tipo bibliográfico, como foi definida no ponto 2.3.1 aponta para um modelo descritivo exigente, semelhante àquele que a teoria biblioteconómica desenvolveu. No entanto, a aplicação” *ipsis verbis*” do modelo tradicional é desajustada, pelo menos para todo o universo, tendo em conta as características destes recursos que circulam em ambiente *Web* (cf. Ponto 2.1), sendo o primeiro problema o da sua identidade. Como afirma Svenonius: “Um documento digital – como um documento hipertextual ou uma mensagem de correio electrónico – pode ser instável, dinâmico, e sem fronteiras identificáveis. (...) Documentos sem fronteiras identificáveis, que estão em evolução, continuamente a crescer, ou a substituir partes deles mesmos têm

problemas de identidade. Não é possível manter a identidade através de um fluxo. (...). Isto não é apenas uma questão filosófica, dado que o que é difícil de identificar é difícil de descrever e portanto difícil de organizar.” (SVENONIUS: 2001, p. 13)

Um outro problema decorrente da questão anterior diz respeito ao conceito de publicação, quer dizer, como indicar as alterações que um objecto digital conhece ao longo da sua existência e que indicadores devemos usar para reconhecer essas alterações. A solução tem passado por designar as modificações com termos como versão, nível, revisão, actualização, que, no entanto, não estão standardizados e portanto não são fiáveis como indicadores de edição. Esta ausência de standardização pode conduzir, segundo Svenonius, a apreciações subjectivas, o que, por sua vez, conduzirá a “povoar o universo bibliográfico de fantasmas e subverter os esforços de standardização.” (SVENONIUS: 2001, p. 40). Ainda neste domínio da publicação, fica pouco claro que informação utilizar para a descrever, por exemplo, o atributo editor.

A descrição de atributos físicos destes recursos constitui uma outra área problemática, tendo-se vindo a verificar uma deslocação de importância dos atributos físicos para as indicações das condições de acesso aos documentos. No entanto, como aduz Svenonius, “Num estudo recente, A. Carlyle mostra que o formato físico é o atributo seleccionado em primeiro lugar quando se pede aos utilizadores para agrupar documentos por características que sejam úteis para efeitos de pesquisa.” (SVENONIUS:2001, p. 110). Estas dificuldades de descrever o invólucro físico da informação poderiam eventualmente pôr em questão a necessidade de uma linguagem para este tipo de atributos. No entanto, embora a função de inventário deste tipo de linguagens perca a sua razão de ser para este tipo de recursos, o documento continuará a ser a espinha dorsal da descrição. Até porque, segundo Svenonius, não existe alternativa: “Informação sem invólucro físico (...) não pode ser identificada nem descrita. Só quando é limitada, no sentido de ser “empacotada” num documento, electrónico ou outro, se torna acessível e só nessa circunstância é candidata à descrição bibliográfica.” (SVENONIUS: 2001, p. 109)

Finalmente, uma última questão problemática assume uma importância fundamental no que à recuperação da informação diz respeito, já que estão em causa os objectivos fundamentais de reunião de conjuntos e de navegação no universo bibliográfico criado. Trata-se do problema do controlo de autoridade, sobretudo no que diz respeito a nomes e assuntos. Como afirma Martin Dillon " O problema de coordenar e diferenciar nomes (...) torna-se um pesadelo no ambiente *Web*. Todos os habituais suspeitos estão envolvidos: nomes individuais, nomes colectivos, nomes geográficos, descritores de assunto; todos compostos actualmente por uma confusão de línguas e conjuntos de caracteres numa imensa escala.” (DILLON: 2000)

O controlo de autoridade, tal como é posto em prática pelas bibliotecas, permite uma qualidade de recuperação porque "O controlo de autoridade permite precisão para recuperar apenas os registos de itens de interesse, e a estrutura sindética das referências cruzadas do controlo de autoridade assegura a recuperação de todos os materiais relevantes, bem como a navegação para chegar a materiais bibliográficos relacionados.” (TILLET: 2001).

No entanto as dificuldades de aplicação do conceito tradicional de cabeçalho uniforme, bem como as dificuldades operacionais e culturais de harmonizar diferentes tradições

nacionais, conduziram recentemente a uma reflexão sobre a aplicabilidade destes princípios da qual ressaltam conclusões que apontam para:

- a) O respeito pelos usos nacionais, mas aproveitando as potencialidades tecnológicas nomeadamente no que diz respeito à implementação da interoperabilidade, caminhar para a construção de um ficheiro de autoridade virtual, que interligue as diferentes semântica e sintaxes de um nome;
- b) Consequentemente a alteração do conceito de cabeçalho uniforme para cabeçalho autorizado, permitindo assim diferentes formas de um nome de acordo com as tradições nacionais. (CAMPOS: 2003)

Entretanto, é preciso ter em linha de conta que existe a montante uma outra questão que é a da existência de um grande número de registos de autoridade autor disponíveis em bases de dados de bibliotecas, muitas disponíveis na própria *Web*. Esta circunstância levanta a questão de como estender a utilização desta informação ao ambiente *Web* e, consequentemente, como assegurar que eventuais registos de autoridade elaborados neste ambiente possam ser utilizados pelas bibliotecas. Barbara Tillett (TILLET:2001) aponta para uma solução possível: a cooperação entre bibliotecas que detêm registos de autoridade e os produtores de metadados, utilizando as funcionalidades do protocolo Z39.50.

A opção por uma linguagem de metadados mais ou menos estruturada, com maior ou menor controlo vocabular tem que ser perspectivada em termos de contexto. Como afirma Lynch “O contexto de utilização é tudo, e só quando falamos de contextos de utilização é que podemos falar de metadados de forma informada e com sentido.” (LYNCH: 2002, p. 183). Quando falamos de contextos de utilização falamos sobretudo de tipos de utilizadores diversos com necessidades diferentes e, portanto, com diferentes expectativas quanto às funções que sistemas de organização da informação devem cumprir, podendo, no entanto, como já referimos (Cf. ponto 2.3.1) considerar-se a existência de objectivos mínimos para estes sistemas. A questão está em saber se no ambiente dos recursos *Web* esses objectivos mínimos têm lugar para todos os tipos de utilizadores e todos os tipos de recursos.

## **2.5. Análise e representação do conteúdo**

O acesso pelo assunto constitui uma das portas de acesso à informação mais utilizadas por todo o tipo de indivíduos e num ambiente onde as dificuldades de identificação da autoria e de determinação exacta de títulos são evidentes torna-se ainda mais relevante.

### **2.5.1. Linguagens de representação**

As linguagens de assunto pretendem representar o conteúdo intelectual da informação, ou seja, dar notícia sobre o que trata uma dada entidade informativa. Segundo Hjorland, os dados sobre o assunto desempenham duas funções:

- a) A função principal é a de fornecer informação sobre o documento que possa auxiliar o utilizador no processo de pesquisa
- b) A função secundária é a de poder substituir o documento original ou representar uma sua condensação (HJORLAND: 1997, p. 27)

Os objectivos essenciais que se pretendem cumprir com a utilização destas linguagens são os de reunião e navegação. Para atingir o primeiro objectivo a linguagem deve ser desenhada de forma a facilitar a recuperação de todos e só todos os documentos pertinentes para a resposta a uma dada questão. Para atingir o segundo objectivo a linguagem deve ser estruturada para mostrar conhecimento relacional.

Em termos de vocabulário a característica principal destas linguagens é a de que se trata de vocabulário controlado, normalizado, podendo esta normalização assumir diferentes níveis, desde a simples enumeração dos termos autorizados até à complexa estrutura de uma classificação. A definição de um vocabulário para estas linguagens envolve, normalmente, o tratamento semântico dos termos para restringir o seu sentido e tornar explícitas as relações que os termos estabelecem entre si. O resultado é um vocabulário ordenado no sentido de que cada termo se refere apenas a um conceito ou objecto e cada conceito ou objecto é designado por um só termo. Os léxicos destas linguagens (classificações e tesouros) são, desta forma, compostos por termos com um sentido especial, ou seja, palavras simples ou combinações de palavras (termos compostos) que constituem o vocabulário técnico de uma dada disciplina do conhecimento. Dentro dessa disciplina são independentes do contexto, no sentido de que podem funcionar sozinhos como indicadores do conteúdo. (SVENONUIS: 2001, pp. 128-129)

São vários os dispositivos de controlo semântico utilizado por estas linguagens. O primeiro diz respeito à classificação do vocabulário em categorias, depois utilizadas para a construção sintáctica, sendo que a definição de categorias não é requisito cumprido por todas estas linguagens. As linguagens que utilizam uma dada sintaxe são designadas por sintéticas, estabelecendo um conjunto de regras simples e finitas, e distinguem-se das que não o fazem que são designadas de enumerativas e nas quais todas as expressões admitidas na linguagem são enumeradas, sendo basicamente o caso da maioria dos tesouros. O controle da ambiguidade terminológica da linguagem natural é feito através do tratamento individualizado de cada sentido dos termos, recorrendo para tal, em muitos casos, à utilização de qualificadores parentéticos, como no caso Perú (ave), Perú (país). Por último, as linguagens de representação do assunto estabelecem, de forma normalizada, as relações entre os termos, basicamente de três tipos: relações hierárquicas, relações entre sinónimos e relações entre termos semanticamente próximos. (SVENONIUS:2001 pp.129-131)

A concretização prática destas linguagens deu origem a diversos instrumentos de organização da informação, que registam algumas diferenças. Podemos agrupá-los em dois grandes tipos: classificações e tesouros.

As classificações são linguagens que estruturam o conhecimento com base num princípio hierárquico, do geral para o particular e fazem-no enumerando, de forma quase sistemática, todas as subdivisões temáticas possíveis de uma dada área do conhecimento. Cada subdivisão é normalmente codificada por uma notação numérica ou alfa-numérica, traduzindo esta codificação a estrutura hierárquica das subdivisões. O critério para estruturar a codificação pode divergir de linguagem para linguagem, por exemplo, na Classificação Decimal Universal utiliza-se a divisão decimal. Nestas linguagens, as relações entre os termos estão definidos à partida, quer dizer trata-se de linguagens pré-coordenadas, onde a relação predominante entre os termos – a hierárquica – é definida à priori. Algumas classificações têm vindo a diminuir a sua

característica enumerativa apostando num desenvolvimento por facetas, como é o caso da C.D.U.

Os tesouros são, ao nível infraestrutural, listas alfabéticas de termos autorizados para nomear conceitos e objectos de uma dada disciplina do conhecimento. Um dos aspectos essenciais e distintivos destas linguagens é que definem diversos tipos de relações entre os termos, a saber: relações de equivalência, relações hierárquicas e relações associativas (CLEVELAND: 2001, p. 43). As relações de equivalência traduzem-se pela remissão de um termo não autorizado para um termo autorizado, podendo assim fornecer um conjunto de sinónimos úteis para criar alternativas à pesquisa e aproximar o vocabulário da linguagem do utilizador. As relações hierárquicas traduzem-se na criação de “árvores” relacionais entre os termos, de acordo com o princípio do geral para o particular, sem, no entanto, estabelecer uma hierárquica com mais de um nível, ou seja, temos apenas termos gerais e termos específicos. As relações associativas traduzem-se na definição de relações semânticas entre os termos quer ao nível dos termos mais gerais, quer dos termos mais específicos. Este tipo de relações são normalmente mais específicas do que numa classificação. Também ao contrário das classificações, não se encontra definida à partida um tipo de relação entre os termos, estes são atribuídos de forma a poderem ser combinados posteriormente para representar conceitos mais específicos, ou seja, estamos perante uma linguagem pós-coordenada.

As diferenças entre estes dois tipos de linguagens foram inicialmente bem delimitadas. Contemporaneamente, no entanto, os “tesouros evoluíram para instrumentos mais complexos, mostrando relações hierárquicas entre os termos e / ou características de uma classificação completamente facetada. Actualmente, os tesouros são utilizados em sistemas pré-coordenados, e os esquemas classificatórios vêm também desempenhando um papel nos sistemas pós-coordenados.” (CLEVELAND: 2001, p. 41). No entanto, é importante referir que as classificações, graças à sua característica mais sintética, proporcionam uma concretização mais eficaz da reunião de conjuntos embora com perda de especificidade na representação, enquanto que os tesouros permitem uma maior eficácia na distinção do que é semelhante.

### **2.5.2. A utilização das linguagens para a representação do assunto no contexto de metadados**

Todas as linguagens de metadados consideram o atributo **assunto** como um dos atributos essenciais da informação, recebendo um tratamento cuidado em termos de definição e utilização. Considera-se a utilização de vários tipos de linguagens, das mais gerais às mais específicas e das mais às menos estruturadas. A utilização destas linguagens documentais para a representação do assunto destina-se a assegurar quer a pesquisa por palavra-chave, quer a organização de estruturas que propiciam o *browsing* das descrições dos recursos.

A reflexão sobre a representação do **assunto** neste contexto e proposta de uma abordagem a esta questão foi levada a cabo com maior sistematicidade, entre 1997 e 1999, por um subcomité designado *Subcommittee on Metadata and Subject Analysis* criado no âmbito do *Subject Analysis Committee (SAC)* da *Association for Library Collections & Technical Services (ALCTS)*. Do trabalho deste subcomité resultou o

relatório “*Subject data in metadata record: recommendations and rationale*”, de 1999. As recomendações deste relatório abrangem todos os aspectos da representação do conteúdo em metadados, tendo a sua origem dentro do modelo de abordagem tradicional à organização da informação, para além de terem sido produzidas por especialistas de reconhecido renome internacional.

Em termos de vocabulário, semântica e sintaxe, os autores deste relatório recomendam:

- a) A combinação de vocabulários controlados e não controlados
- b) A utilização de múltiplos vocabulários: classificações, tesouros; vocabulários generalistas e/ ou específicos
- c) Os sinónimos devem ser tratados através do design da implementação do vocabulário controlado, ou em alternativa através da inserção na descrição de todos os sinónimos e termos relacionados
- d) A sintaxe do atributo deve ser flexível, mais ou menos complexa conforme a experiência e formação dos indexadores
- e) O desenvolvimento e refinamento de métodos de harmonização entre termos de diferentes vocabulários deve ser levado a cabo
- f) A utilização de várias classificações deve ser permitida, mesmo dentro de um caso particular. Devem ser permitidas várias notações na mesma descrição, desde que convenientemente atribuídas e codificadas.

A aplicação dos vocabulários deve ser norteadada pelas seguintes recomendações:

- a) Adopção de vocabulários existentes, com ou sem adaptações, tendo em conta a necessidade de assegurar a interoperabilidade semântica
- b) A selecção dos vocabulários apropriados deve considerar os seguintes critérios: assuntos que o vocabulário representa, o público-alvo, nível de especificidade na representação, a natureza e abrangência da colecção de recursos, e a sintaxe
- c) Cada organização deve estabelecer uma política sobre o nível apropriado de representação do assunto para a sua colecção e a este nível deve optar-se pelo nível de representação mais específico.
- d) Dentro de um determinado projecto, semântica e sintaxe devem ser aplicadas de forma consistente
- e) A versão do vocabulário utilizado deve ser indicada
- f) A classificação deve ser utilizada a um nível específico
- g) A notação da classificação deve ser incluída na descrição, sendo desnecessário a inclusão dos descritores associados. Se tal for necessário deve ser resolvido pelo *design* do sistema

Das recomendações deste relatório ressaltam como mais significativos os seguintes aspectos: a possibilidade de utilizar diferentes tipos de vocabulário controlado, adaptado ou não; a utilização de diferentes vocabulários do mesmo tipo, por exemplo classificações; a representação deve ser realizada ao maior nível de especificidade possível; a opção por utilizar as notações das classificações e não os descritores directamente; a necessidade de harmonizar diferentes vocabulários numa implementação; a consideração de critérios objectivos para a selecção dos vocabulários mais apropriados e, finalmente, a necessidade de estabelecer uma política de classificação e indexação em cada implementação.

Se comparadas estas recomendações com as estabelecidas pelo *DESIRE information gateways handbook* (ver ponto 3.2.1) verifica-se uma quase completa concordância em todos os aspectos significativos, exceptuando a possibilidade de utilizar vocabulários não controlados.

### **2.5.3. *Browsing* de recursos**

A organização de recursos de informação através de uma estrutura em níveis baseada na selecção de determinados termos gerais e específicos tem vindo a ser implementada por praticamente todos os serviços de acesso à *Web*. Estas estruturas, graças as potencialidades do hipertexto, permitem a fácil navegação entre os termos. Assim, podemos definir estas formas de organizar o acesso à informação, normalmente em forma de árvore, como uma colecção de assuntos organizados hierarquicamente sob os quais são listadas descrições de conjuntos de recursos, facilmente navegáveis (WHEATLEY: 2000) O termo navegação, usado para indicar um movimento através de um espaço de pesquisa num interface gráfico é limitado para um completo entendimento e caracterização destas estruturas, bem como relativamente às funções que podem desempenhar enquanto formas de organização e acesso à informação.

O conceito de *browsing*, nas dimensões comportamental e de organização dos recursos adequada ao exercício desses comportamentos, permite equacionar melhor os dois principais aspectos envolvidos nesta realidade: por um lado, os indivíduos e os seus comportamentos de pesquisa de informação, e por outro, os processos de organização da informação. No âmbito deste trabalho, a perspectiva mais importante é a que diz respeito aos processos de organização da informação, mas estes só podem ser completamente compreendidos se tomarmos em linha de conta que eles pretendem ou devem responder a determinados comportamentos humanos de pesquisa de informação.

Do ponto de vista comportamental, o conceito de *browsing* é de difícil definição por querer exactamente traduzir, pelo menos numa das acepções possíveis, um comportamento de pesquisa muito pouco estruturado. Numa primeira acepção, *browsing* pode ser considerado como a probabilidade de encontros acidentais com recursos de informação, que podem resultar de um interesse pré-existente ou causar um novo interesse (CASE:2002). Trata-se então de uma actividade não orientada, não programada, informal e não sistemática (NAVES:1998).

A investigação sobre comportamentos de pesquisa de informação tem vindo a alargar este conceito incluindo um comportamento de pesquisa mais direccionada e distinguindo então diversos níveis de *browsing*, que recobrem um conjunto diversificado de comportamentos entre o extremo que vai da descoberta acidental de informação até à pesquisa estruturada em função de objectivos bem definidos, podendo passar por níveis de semi-definição.

Aplicando esta perspectiva à descoberta de recursos no ambiente que estamos a estudar um comportamento de *browsing* bem definido traduz-se em procurar recursos específicos ou registos utilizando termos controlados ou atributos da informação, um comportamento semi-definido em encontrar recursos coincidentes com termos gerais de linguagem natural, um comportamento pobremente definido o de seguir ligações entre

páginas consideradas interessantes e, finalmente, um comportamento indefinido o de acidentalmente encontrar páginas de interesse (CASE:2002, p.87).

Este alargamento conceptual da noção de *browsing* parece pecar por excesso, na medida em que então todo o comportamento de pesquisa pode ser denotado por este conceito, perdendo assim a sua capacidade discriminadora que o permitiria distinguir de outros tipos de comportamento, que podem efectivamente ser exercidos através de outras formas de pesquisa, nomeadamente a pesquisa por palavra, a qual, se bem que nela também possa ter lugar um comportamento pouco definido, tem condições específicas para uma pesquisa mais estruturada, nomeadamente aquela que recorre aos atributos da informação, intrínsecos ou extrínsecos.

Por outro lado, este conceito tão alargado de *browsing* não é operativo para a análise da realidade objecto de estudo, sobretudo no extremo a jusante, ou seja, o do comportamento direccionado que se traduz na pesquisa por termos controlados ou atributos dos recursos. De facto, este tipo de comportamento só pode ser exercido neste contexto por via da pesquisa por palavra, que para efeitos de performance de recuperação apresenta níveis similares aos proporcionados pelo *browsing* (WHEATLEY:2000).

Assim, no contexto em estudo podem enquadrar-se basicamente dois níveis de *browsing*. Um primeiro, que podemos qualificar de bem definido, em que o utilizador define claramente o seu objectivo, em termos de assunto por exemplo, e percorre a estrutura que lhe é disponibilizada à procura do tema seleccionado. Um segundo, que poderemos qualificar de semi-definido e no qual o utilizador, que aborda qualquer um destes serviços com alguma indicação do que pretende pesquisar dada a especificidade temática de todos eles, não tem um objectivo específico previamente definido, e percorre a estrutura organizadora dos recursos para ter uma visão geral dos conteúdos disponíveis. A organização dos recursos deve, então, permitir o exercício destes dois níveis de *browsing*.

A eficácia de qualquer organização para efeitos de *browsing* depende ainda das garantias que pode fornecer quanto à consistência na representação da informação e da possibilidade reunir um conjunto de recursos que partilhem qualquer instanciação do ponto de acesso seleccionado. Assegurar o primeiro destes requisitos implica o recurso a vocabulários controlados, enquanto que o segundo pode ser implementado com recurso a várias técnicas, desde o controlo de autoridade até à ligação transparente entre termos alternativos.

A selecção do ponto de acesso que será a base da estrutura de *browsing* parece não oferecer grandes dúvidas: trata-se quase sempre do assunto no sentido do termo ou termos que extraídos de uma dada linguagem controlada representam o conteúdo do documento. A selecção deste ponto de acesso prende-se claramente com a natureza do comportamento de *browsing*: mesmo nos casos em que ele possa ter um objectivo mais definido, *browsing* é sempre uma actividade pouco estruturada, pelo que é mais significativo para o utilizador aceder a um conjunto de assuntos organizados de uma dada forma, que a uma lista estruturada de autores ou de títulos.

A organização de estruturas de *browsing* para acesso por **assunto** com base em vocabulários controlados fez renascer uma importância não prevista para as linguagens

documentais tradicionais. Em primeiro lugar, porque a utilização de vocabulários controlados assegura consistência na representação do conteúdo, precisão na recuperação e controlo na atribuição. Em segundo lugar, porque a longa história destes instrumentos de representação proporcionou a existência de uma grande diversidade de tipos, desde os mais gerais aos mais específicas, desde os mais hierarquizados aos mais relacionais. Em terceiro lugar, a larga utilização pelas bibliotecas destes instrumentos criou as competências necessárias nos profissionais para a sua utilização e tornou-os familiares para um grande número de indivíduos. O baixo custo de produção e facilidade de aplicação particularmente bem adaptados a colecções mais voláteis de recursos, sobretudo se o vocabulário não estiver muito desenvolvido, e a simplicidade de acesso para utilizadores pouco experientes são outros dois aspectos que contribuem para uma cada vez maior utilização de vocabulários controlados para a implementação deste paradigma de *browsing*.

De entre os dois grandes tipos de linguagens documentais existentes, as classificações, que basicamente se traduzem pela aplicação do princípio básico do agrupamento hierárquico pela similitude dos assuntos, apresentam um conjunto de vantagens específicas, algumas podendo ser particularmente bem adaptadas a este ambiente. A sua estrutura permite uma constante contextualização dos assuntos ao longo de todos os passos da navegação, alargar ou restringir a pesquisa de forma ágil, identificar os conteúdos a níveis de especificidade diferenciados. Na medida em que a notação é independente da língua em que o descritor é referenciado, a sua utilização permite o acesso a utilizadores experientes que não dominem a língua utilizada. Graças à ausência de fisicalidade estrita do recurso e às potencialidades do hipertexto, é possível atribuir o mesmo recurso a diferentes classes, bem como estabelecer com facilidade referências cruzadas entre as várias classes ou sub-classes. Se bem exploradas, estas possibilidades permitem ultrapassar uma das principais desvantagens das classificações e que é a sua rigidez na contextualização dos assuntos numa determinada classe, o que conduz, qualquer que seja o critério utilizado, a uma certa ilogicidade, em muitos casos, na escolha da classe onde será integrado um determinado assunto.

À lista de vantagens identificadas, Diane Vizine-Goetz acrescenta ainda: a capacidade de cobertura de grande variedade de assuntos, a existência de níveis de generalidade e especificidade suficientes para suportar o *browsing*, e finalmente, as notações podem ser utilizadas para manipular com facilidade as categorias quer para efeito de *browsing* quer de pesquisa por “palavra”. (VIZINE-GOETZ:2002)

Para além da desvantagem já referida, a utilização de classificações apresenta outras, nomeadamente: a lentidão na representação de novos conceitos, a dificuldade em representar relações diversas entre conceitos.

A utilização de tesouros neste contexto apresenta a grande vantagem de permitir uma representação do conteúdo dos documentos a um nível de muito maior especificidade do que o das classificações. O objectivo da utilização de uma linguagem deste tipo não é tanto o de associar documentos que pertencem a uma mesma classe, mas o de representar com alta especificidade os conteúdos dos documentos individuais. Este tipo de linguagem aponta predominantemente para o exercício da coordenação entre os termos no momento da pesquisa (pós-coordenação), sendo, por isso, menos adequada para a criação de árvores hierárquicas de assuntos. Como afirma Schwartz “É difícil usar a indexação pós-coordenada como mecanismo de *browsing*, dado que cada termo

pode indexar um largo número de itens e a informação contextual está ausente.” (SCHWARTZ: 2001, p.90) No entanto, as relações entre os termos num tesauros podem permitir a criação de uma estrutura que as traduza, embora a multiplicidade de relações possíveis dentro de cada micro-tesauros e entre estes torne uma tal estrutura muito complexa para ser facilmente percebida pelo utilizador.

O nível de especificidade que é possível atingir com uma linguagem deste tipo torna-a particularmente adaptada para situações em que o utilizador já conhece o termo ou termos que pretende pesquisar.

#### **2.5.4. Descrição em texto livre**

Um resumo é uma representação sucinta, mas exacta, do conteúdo de um documento ou recurso de informação, produzido pelo indexador utilizando sobretudo a sua própria linguagem, ainda que possa utilizar citações do documento original. (LANCASTER: 1991).

No contexto de metadados esta noção equivale, normalmente, ao elemento Descrição, se bem que em alguns casos este elemento possa denotar um conceito que se pretende mais alargado. Tal é o caso de Dublin Core, onde este elemento é definido como uma descrição do conteúdo do recurso, o que inclui o resumo, mas também outras formas de representação possíveis tais como a tábua de matérias, referência a qualquer representação gráfica do conteúdo ou a qualquer texto descritivo do recurso, ou outras que se enquadrem neste contexto. (DCES:2003<sup>11</sup>). O carácter mais generalista da noção de Descrição face à de Resumo é reforçado através da definição do qualificador Resumo como um dos possíveis refinamentos do elemento Descrição, definido agora como um sumário do conteúdo. (DCMI metadata terms: 2004). Esta distinção encontra apenas uma justificação de tipo operacional no sentido em que permite distinguir o Resumo, enquanto texto produzido pelo indexador, da mera transcrição de tábua de matérias de um dado recurso, possibilidade esta que se afasta de qualquer uma das tipologias de conteúdos de qualquer tipo de resumo.

Do ponto de vista do conceito, esta distinção encontra pouca justificação na medida em que a noção de resumo é suficientemente lata para incluir a possibilidade de referenciar todos os aspectos considerados significativos no que diz respeito à representação do seu conteúdo, incluindo portanto a tábua de matérias, a possível representação gráfica do conteúdo, etc., desde que mantenham as características básicas de síntese e de meta texto, no sentido em que se trata de um texto sobre outro texto, produzido ou não pelo autor do texto.

Nos casos que estamos a analisar, a definição do elemento descrição é, como teremos oportunidade de ver com detalhe, em tudo semelhante à de um resumo, não tendo nenhum dos serviços sentido a necessidade de definir um qualquer qualificador para o elemento.

A função primordial de um resumo é a de apoiar o utilizador na selecção dos itens adequados à satisfação da sua necessidade de informação. Alguns tipos de resumos,

---

<sup>11</sup> Dublin Core metadata element set: version 1.3: reference description

nomeadamente o designado por informativo, podem cumulativamente desempenhar, se bem executados, a função de substituto do próprio recurso. Nos casos em que o resumo é directamente pesquisável, contribui igualmente para a concretização do objectivo de encontrar um dado recurso, embora sem as potencialidades de discriminação proporcionadas por elementos como o título ou o autor.

Existe na literatura alguma diversidade quanto às tipologias de resumos (LANCASTER:1991, WHEATLEY:1997, ROWLEY:2000), algumas reveladoras de grande imprecisão no que diz respeito ao seu conteúdo. No entanto, podemos identificar três tipos comuns que constituem os mais significativos: o resumo indicativo, o resumo informativo e o resumo crítico. O resumo indicativo limita-se a descrever, indicar o assunto de que trata o documento, enquanto que o resumo informativo vai mais longe e apresenta também os vários aspectos da produção do conhecimento de que o documento em causa trata. No caso concreto de um relatório de uma investigação, o resumo indicativo limitar-se-á a sinalizar o tipo de resultados finais, enquanto que o resumo informativo sintetizará esses resultados e informará sobre as metodologias aplicadas. Alguns consideram que este tipo de resumo é particularmente adaptado a documentos que versem um único tema ou transmitam o resultado de uma investigação (ROWLEY:2000). O resumo crítico tem um carácter explícito de avaliação do conteúdo, sendo mesmo definido como uma “recensão crítica condensada” (LANCASTER:1991, 90). Enquanto que para os dois primeiros casos, não é indispensável que o indexador seja alguém com conhecimentos específicos no assunto abordado, já para o último tal não se verifica. Por isso, os resumos verdadeiramente críticos são raros.

Apesar do esforço de definição e distinção entre as várias tipologias, a prática tem vindo a demonstrar a dificuldade em produzir categorias “puras”. Assim, a realidade mais comum é a de resumos mistos que juntam características das tipologias informativo e indicativo, designados por resumos indicativos-informativos, nos quais os elementos essenciais do conteúdo são tratados de acordo com a tipologia informativa e os elementos secundários de acordo com a tipologia indicativa (ROWLEY:2000). A selecção do tipo de resumo mais adequado a cada caso deve ter em conta certas variáveis: o público-alvo, os recursos humanos e financeiros (já que por exemplo um resumo informativo, por ser normalmente mais longo, exige mais tempo), os objectivos do serviço produtor e a área disciplinar em que se insere o conteúdo. Esta relação entre a área disciplinar do conteúdo e o tipo de resumo é assumida claramente por alguns, que apontam uma maior adequação do resumo informativo para os domínios das ciências exactas, enquanto que os resumos indicativos seriam mais apropriados para as ciências humanas (LANCASTER:1991).

Independentemente da tipologia, os parâmetros extensão e estrutura do resumo, comuns a todas, são importantes elementos caracterizadores destas entidades de informação. São frequentes, entre os serviços produtores de resumos, as indicações relativamente à extensão máxima (em número de caracteres ou de palavras) de um resumo, embora não exista qualquer razão para que todos os resumos tenham a mesma extensão. A extensão de um resumo deve ser vista como o resultado da verificação de diversas variáveis, a saber: extensão do recurso original, a complexidade e diversidade do conteúdo, a importância do item no contexto descritivo, custo e a finalidade com que é feito o resumo (LANCASTER: 1991).

As opções relativamente à estrutura do resumo oscilam entre textos exclusivamente narrativos e resumos estruturados, ou seja, o resumo obedece a uma determinada estrutura temática, como por exemplo problema, análise do caso, metodologia, conclusões (WHEATLEY:1997). O principal problema de um resumo assim estruturado é o de que as categorias só se aplicam num contexto relativamente homogéneo de fontes de informação. Assim, entre o extremo de um texto absolutamente narrativo e organizado de acordo com as idiossincrasias de cada indexador e um estrutura rígida, podem encontrar-se algumas indicações que visam manter o texto narrativo em determinados limites: é o caso de algumas regras relativamente ao número de parágrafos aconselhados ou mesmo permitidos ou a ausência de listagens enumerativas.

Tendo embora em conta a diversidade de possibilidades até aqui analisadas, é possível definir as características de um bom resumo, a saber: brevidade, clareza, rigor, boa legibilidade, suficiente exaustividade na representação do conteúdo e finalmente, senão mesmo em primeiro lugar, a capacidade de servir as necessidades do utilizador (SOON:2002).

No contexto de análise em que se situa este trabalho, a produção de resumos por humanos, ou mesmo com recurso a processos automatizados, apresenta algumas especificidades que entroncam quer na nova forma de produzir a informação – hipertextual – quer no modelo editorial diferente, quer ainda numa muito mais diversa tipologia de recursos informativos. O carácter dinâmico da informação pode implicar que aquilo que o indexador teve oportunidade de observar e sintetizar já não seja exactamente a mesma realidade amanhã. A diversidade de caminhos possíveis graças à hipertextualidade pode conduzir o utilizador a uma experiência de percepção do recurso completamente diferente da do indexador. Os diferentes níveis de granularidade que é possível implementar na descrição, implicam que esta tanto possa ter como objecto um recurso informativo individual como uma colecção de recursos. Ora, as colecções variam ao longo do tempo e a abordagem que o utilizador faz a essa colecção pode ser diferente da do indexador. Estes aspectos levam Wheatley a pugnar pela necessidade de uma taxinomia de resumos mais rica (WHEATLEY:1997). Embora não elaborando uma nova taxinomia, o mesmo autor propõe algumas características mais específicas para os resumos de documentos que circulam em ambiente *Web*, a saber: informação para guiar o utilizador na exploração do recurso, devem proceder a uma avaliação da autoridade, discutir aspectos relativos a atributos físicos, fornecer julgamento sobre a qualidade do conteúdo e, finalmente, fazer referências para fontes alternativas (WHEATLEY:1997, 51).

Esta proposta, até pela data da sua produção, ignora que alguns dos aspectos que são referidos são objecto de tratamento no contexto de estruturas de metadados que não se limitam à descrição do recurso. O caso mais evidente é o relativo à avaliação da autoria: de facto; este problema é claramente resolvido pelo tratamento do elemento responsabilidade que, na grande maioria das estruturas de metadados, e normalmente através de qualificadores vários, permite ao utilizador verificar a validade dos responsáveis para produzir a informação que é disponibilizada. No entanto, não podemos deixar de concordar que uma nova abordagem é necessária para resumos que têm como objecto documentos com características muito diferentes dos que até há pouco tempo existiam e para os quais todos os aspectos que temos vindo a referir foram originalmente pensados.

Tendo em conta a problemática enunciada, e por outro, a facilidade com que é possível aceder ao recurso propriamente dito a necessidade de resumos tornar-se-á menos importante neste ambiente ? Aparentemente a resposta é não. Estudos recentes mostram que, apesar da facilidade de acesso, os resumos produzidos por humanos continuam a desempenhar uma importante função no que diz respeito ao apoio ao utilizador na selecção dos recursos informativos mais adequados às suas necessidades e características (PINTO: 1999).

## Cap. III – Os Gateways no âmbito da RDN (Resource Discovery Network)

### 3.1. O nascimento da ideia e suas concretizações

Os Gateways que formaram, mais tarde a RDN, surgiram como resposta ao desafio que resultou do Relatório Follet.

#### 3.1.1. O projecto eLib e o surgimento dos primeiros Subject Gateways

O projecto eLib (Electronic Libraries) nasce no início da década de 90 como resultado das recomendações do relatório Follett<sup>12</sup>, que se destinava a investigar as formas possíveis de lidar com as pressões sobre os recursos das bibliotecas universitárias causadas pela rápida expansão do número de alunos e pela acelerada explosão do conhecimento e da informação académicas. Uma das principais conclusões deste relatório enfatizava a importância da exploração das Tecnologias da Informação para criar um efectivo serviço de biblioteca no futuro.

Como resposta às recomendações do relatório Follett, os *Higher Education Funding Councils for England, Scotland and Wales* lançam um conjunto de candidaturas para projectos que visem transformar o armazenamento e a utilização da informação e do conhecimento nas instituições universitárias. Nasce, desta forma, o programa eLib que será gerido pelo *JISC (Joint Information System Committee)* entretanto constituído como resultado das directrizes dos Secretários de Estado que incentivavam os recentes organismos de gestão do ensino superior a construírem um comité para tratar dos serviços de informação em rede especializados.

A partir dos primeiros “concursos” dirigidos às instituições universitárias em 1994, é apresentado um significativo conjunto de projectos em várias áreas definidas pela Agenda do JISC e cujo objectivo geral era o da criação, de forma cooperativa, de uma infra-estrutura de informação académica. Em 1995 nascem, então, os primeiros projectos na área do acesso a recursos em rede (*ANR= access to network resources*). Com a constituição desta área específica no âmbito da Agenda do JISC pretendia-se aumentar o conhecimento sobre os recursos em rede, explorar os vários aspectos relacionados com a gestão de serviços em larga escala, assegurar o envolvimento da comunidade universitária no desenvolvimento a nível nacional e internacional, e desenvolver iniciativas com o objectivo de criar uma infra-estrutura nacional capaz de gerar uma utilização significativamente mais alargada de recursos de informação em rede. Para assegurar a consecução destes objectivos foram seleccionadas várias áreas em condições de os concretizar, incluindo a definição de um centro nacional que pudesse actuar como ponto de acesso global e como conselheiro para as questões relacionadas com a informação em rede; promoveu-se o estabelecimento de serviços de acesso à informação pelo assunto (*Subject Services*) para testar problemas de escala associados à disponibilização de um serviço mais alargado, a disseminação da informação sobre sistemas de descoberta de recursos, a cooperação internacional e o desenvolvimento de directrizes e standards. (DEMPSEY:2000)

---

<sup>12</sup> Disponível em <http://www.ukoln.ac.uk/services/papers/follett/report/ch1.html>

É exactamente neste contexto que vão nascer os primeiros *Subject Gateways* da responsabilidade, na grande maioria dos casos, de bibliotecas universitárias: OMNI para as áreas da medicina e biociências, ADAM para as artes e *design*, EEVL para engenharia e RUDI para o *design* urbano. Estes serviços constituem projectos novos que se traduzem, em primeira instância, pela constituição de bases de dados de recursos Internet nas respectivas áreas temáticas. A constituição destes novos serviços beneficiou, sem dúvida, da experiência de SOSIG, *Subject Gateway* criado anteriormente no âmbito do programa DESIRE para a área das ciências sociais. A experiência SOSIG serviu, assim, como paradigma para a criação destes serviços.

A constituição dos *Subject Gateways* foi ainda apoiada dentro do programa eLib através do financiamento do projecto ROADS (*Resource Organization and Discovery in Subject-based Services*), cujo objectivo era o da criação de um pacote de software que permitisse a construção de uma base de dados de descrição dos recursos e um interface *Web* para a sua disponibilização, para além da consideração das questões relativas à interoperabilidade entre diferentes sistemas e ao desenvolvimento de standards para a catalogação, classificação, indexação e pesquisa de recursos.

Ao longo dos anos de 1995 – 1997 foram sendo criados e disponibilizados ao público os primeiros *Subject Gateways*.

### **3.1.2. A RDN (Resource Discovery Network)**

A RDN (*Resource Discovery Network*) foi fundada pelo JISC (*Joint Information Systems Committee*) (Cf. Circular 10/98: *Subject Information Gateways*, do JISC) e lançada publicamente em Novembro de 1999, com os objectivos de alargar a cobertura a áreas temáticas não cobertas pelos *Gateways* criados no contexto do eLib, integrar o acesso à informação a partir dos *Gateways* existentes e dos que viessem a ser criados, desenvolver Portais de assunto para comunidades educativas e, finalmente, estabelecer um novo modelo organizacional e de negócio. Esta rede conta com o apoio do *Economic and Social Research Council* (ESRC) e do *Arts and Humanities Research Board* (AHRB)

A coordenação desta rede é assegurada pela DNER (*Distributed National Resource*) e tem como tarefas: promover e desenvolver a rede, estabelecer áreas de trabalho que visem assegurar a qualidade, consistência e interoperabilidade na RDN, promover de várias formas os *Gateways* explorando a sua interdisciplinaridade e o seu valor.

Um dos desenvolvimentos neste contexto foi a criação do conceito de *Hub*, um desenvolvimento dos *Subject Gateways*, que se traduz essencialmente na tentativa de expandir a cobertura temática de cada um dos serviços adicionando novos *Gateways* ou promovendo a reunião de vários *Gateways* entre si. Trata-se de maximizar o potencial existente para parcerias estratégicas, garantindo desta forma a sustentabilidade e alargando o espectro de público-alvo (DEMPSEY:2000).

Os primeiros *hubs* a serem criados foram os seguintes: SOSIG (*Social sciences, business & law*) PSIGATE (*Physical sciences*), HUMBUL (*Humanities*), BIOME (*Health & life sciences*) e EEVL (*Engineering, mathematics & computing*).

Recentemente três novos serviços foram disponibilizados: ALTIS (*Hospitality, sport, tourism and leisure*), Gesource (*Geography & environment*) e Artifact (*Arts & creative industries* – Novembro de 2003).

A RDN disponibiliza um serviço de pesquisa cruzada por palavra-chave em todos os *Gateways*, podendo o utilizador pesquisar as descrições produzidas por qualquer um dos serviços a partir do mesmo interface.

A funcionalidade de *browsing* dos recursos está disponível apenas através de uma hiperligação para cada estrutura dos serviços, não existindo, portanto, qualquer possibilidade de navegação cruzada.

No sentido de assegurar a interoperabilidade entre os vários *hubs*, até para garantir a possibilidade de pesquisa cruzada, a RDN desenvolveu um conjunto de regras da descrição (DAY:2003) que pretendem ser o mínimo denominador comum das várias regras dos serviços e que contemplam o seguinte conjunto mínimo de atributos: **título**, assunto, descrição, tipo, identificador e língua. Foi igualmente definida uma lista de tipos de recursos a ser utilizada pelos *Gateways*.

No que respeita à gestão das colecções foi também desenvolvido um documento (JENNINGS:2002) orientador para harmonizar os processos de gestão e desenvolvimento das colecções nos vários serviços.

No âmbito da RDN têm vindo a ser desenvolvidos pacotes de formação online que visam desenvolver as competências de acesso à informação na Internet para a comunidade universitária. Actualmente existem 61 tutoriais temáticos, baseados no princípio da auto-aprendizagem<sup>13</sup>.

Em Setembro de 2001 iniciou-se o projecto SPP (*Subject Portals Project*)<sup>14</sup>, cuja primeira fase terminou em Agosto de 2003 e cujo objectivo era transformar os existentes *hubs* em Portais. O conceito de Portal aqui presente traduz-se pela existência de um *Subject Gateway* que, para além do catálogo de recursos individual, disponibiliza através de um único ponto de entrada o acesso a informação com origem em várias fontes, bem como serviços de personalização e partilha da informação dentro da comunidade de utilizadores. (CLARK:2001). No decurso desta fase verificou-se uma redefinição dos objectivos no sentido não de criar exactamente portais mas funções e serviços típicos de portais que os *Hubs* iriam integrando progressivamente, nomeadamente: pesquisa cruzada de recursos de informação de vários editores, serviço de alerta, serviço agregado de notícias e, por último, serviços adicionais que, conforme o interesse das comunidades de utilizadores, podem revestir a forma de motor de pesquisa para jornais electrónicos ou listas de anúncios de emprego entre outros (MARTIN: 2003). A segunda fase deste projecto, que se desenvolve entre Setembro de 2003 e Agosto de 2004, pretende estender este tipo de serviços e funcionalidades a todos os *hubs* da RDN. (Sítio da RDN: 2003).

---

<sup>13</sup> Disponíveis através de RDN Virtual Training Suite em <http://www.vts.rdn.ac.uk>

<sup>14</sup> Informação disponível em <http://www.portal.ac.uk/spp>

## **3.2. Paradigmas e instrumentos de organização da informação para *Subject Gateways***

Nas opções de organização da informação, estes serviços foram influenciados, por um lado, por uma conceptualização teórica do que deveria ser um *Subject Gateway*, corporizada pelo *DESIRE Information Gateway Handbook*, e de ponto de vista prático por uma aplicação que pretendia apoiar essa conceptualização à qual se associaram igualmente Regras de Catalogação, corporizada no pacote aplicacional *ROADS* (*Resource Organization and Discovery in Subject- Based Services*)

### **3.2.1 O “paradigma” DESIRE (Desire Information Gateways Handbook)**

O Projecto *DESIRE* (*Development of a European Service for Information on Research and Education*) foi desenvolvido entre 1998 e 2000, apoiado pela EU, por instituições, maioritariamente bibliotecas holandesas, norueguesas, suecas e inglesas. O objectivo principal era desenvolver as redes de informação já existentes na Europa para um público-alvo de investigadores e universitários, através de investigação nas seguintes áreas: Descoberta de Recursos e Serviços de Directório. Neste último domínio, uma das principais linhas de investigação foram os *Subject Gateways*, da qual resultou uma proposta de organização e gestão desses serviços que passaremos a analisar.

Um dos principais resultados da linha de investigação referida foi a edição electrónica de um manual<sup>15</sup> (1ª versão de 1999) de apoio para todas as organizações, bibliotecas ou outras, que tivessem por objecto a criação de um *Subject Gateway*. O manual aborda as principais questões que se levantam na construção e manutenção deste tipo de serviços, a saber: questões relacionadas com a gestão do serviço, com a organização e recuperação da informação e, finalmente, as relativas aos aspectos tecnológicos. A adopção pelos *Gateways* de práticas estandardizadas permitirá operacionalizar a pesquisa cruzada entre diversos serviços.

Um dos principais aspectos distintivos da abordagem dos *Subject Gateways* face aos recursos *Web* que seleccionam é o de constituírem um verdadeiro catálogo de recursos. Esta abordagem traduz-se na elaboração de descrições desses recursos, por técnicos especializados em descrição, através da utilização de uma dada linguagem de metadados. As descrições são então armazenadas numa base de dados, tipicamente de tipo relacional. A adopção desta abordagem tem, segundo o manual que vimos a seguir, as seguintes vantagens:

- a) Os metadados têm uma dada estrutura e por isso permitem facilidades de pesquisa mais avançadas (por exemplo, pesquisa por campo, como o autor ou o título)
- b) Os metadados podem ser convertidos noutros formatos ou utilizados para garantir a interoperabilidade com diferentes protocolos de pesquisa e recuperação
- c) É mais fácil manter um base de dados de descrições de recursos do que um grande número de ficheiros HTML. Metadados de administração podem também ser utilizados para registar as circunstâncias segundo as quais um recurso necessita de ser reavaliado ou removido da base de dados.

---

<sup>15</sup> A análise aqui realizada baseia-se *Desire information gateways handbook*. Disponível em: <http://www.desire.org/handbook>

Alertando para a necessidade de os responsáveis pelo *Subject Gateway* identificarem os atributos da informação que pretendem descrever, o Manual avança um conjunto de atributos divididos de acordo com a seguinte tipologia: metadados descritivos, metadados para representação dos assuntos e metadados para a administração da descrição e do serviço.

No primeiro caso, os metadados descritivos, são apontados os seguintes atributos: **título, título abreviado** (isto é, um acrónimo do título completo), **título alternativo** (isto é, título do recurso em outra língua), **complemento de título, descrição, URI** (ou outra localização), **autor, língua, codificação do conjunto de caracteres, organização** – criadora ou detentora do recurso, **media** – texto, imagens, áudio, vídeo, **tipo de recurso, media físico, detentor dos direitos de autor, disponibilidade** (por exemplo, pagamento ou registo necessário), **software requerido para acesso, avaliação da qualidade** (*quality rating*), **público alvo**.

Para a área da representação do assunto, aconselham-se os seguintes atributos: **palavras-chave, código de classificação, sistema de classificação** – deve acompanhar o código de classificação, **termos de tesouros, cabeçalhos de assunto**.

Neste conjunto de atributos sugeridos (23) predominam claramente os que dizem respeito à **Obra** enquanto entidade intelectual, a saber: **título, título abreviado, subtítulo, descrição, autor, língua, organização, tipo de recurso, avaliação da qualidade, público-alvo**, aos quais se adicionam todos os relativos ao **assunto**. Os atributos relativos ao **Documento** dizem genericamente respeito aos seguintes aspectos: localização, formato, acesso e propriedade intelectual. Apesar desta predominância da entidade **Obra** enquanto objecto da descrição, a noção de recurso não é claramente definida podendo apenas concluir-se que a cada descrição equivale um identificador na rede (URI), ou seja, os recursos são vistos enquanto DLO's (*Document like object*). Esta relativa indefinição acarreta uma outra, que é a de saber qual o nível de granularidade ou detalhe a utilizar na descrição de cada recurso, produzindo previsivelmente descrições de objectos muito diferenciados, verificando-se ora a descrição de um sítio *Web* na sua totalidade ora a descrição de uma das partes, por ser eventualmente a mais pertinente tendo em conta a temática do serviço e o público alvo.

No que diz respeito ao vocabulário a utilizar para designar os atributos seleccionados, os autores do manual elaboram recomendações relativamente à produção de Regras de Catalogação para cada *Gateway* (cf. Capítulo 2.4. do Manual) que denotam uma preocupação importante em assegurar uma normalização interna e um controlo vocabular relativamente profundo. Assim, de acordo com o Manual, as Regras de Catalogação de cada serviço devem estabelecer neste domínio: uma breve explanação sobre que tipo de informação cada elemento deve conter, directrizes quanto à utilização de formatos para conteúdos como datas, códigos de línguas, nomes, etc. Chama-se igualmente a atenção para a necessidade de estabelecer regras de forma relativamente a nomes e títulos alertando-se para a complexidade do problema. O controlo de autoridade dos nomes é deixado ao critério de cada serviço, indicando-se, no entanto, o recurso a ficheiros de autoridade já existentes. No domínio da Sintaxe, indica-se a necessidade de estabelecer regras relativamente à forma de entrada da informação em cada atributo definido.

A importância do acesso por assunto neste tipo de serviços conduz os autores do Manual ao desenvolvimento deste aspecto, para além dos atributos já identificados. A sua reflexão e aconselhamento distribuem-se pelas seguintes áreas: utilização de classificações, utilização de palavras-chave e tesouros, funcionalidades de *browsing* e pesquisa.

Depois de elaborarem um elenco de vantagens e desvantagens na utilização das classificações e de analisarem comparativamente os diversos tipos de esquemas de classificação, sugerindo claramente a utilização de um esquema já existente, os autores abordam dois aspectos que se prendem mais com a implementação e adaptação ao tipo de recursos e serviços, constituindo, particularmente o segundo, o aspecto mais importante na medida em que se trata de adaptar a tradição ao novo ambiente.

A adaptação das classificações às funções que se pretende que desempenhem neste contexto pode assumir duas formas: adaptação propriamente dita e extensão. No primeiro caso, prevêem-se as seguintes possibilidades: omitir classes vazias, reorganizar as hierarquias e renomear os descritores associados às notações. Uma desigual distribuição dos recursos pelas classes pode gerar confusão aos utilizadores, devendo as classes ser omitidas completamente quando for grande o número de áreas vazias ou claramente marcadas como vazias nos casos em que essa quantidade não seja particularmente significativa. No entanto, prevendo o crescimento da estrutura de *browsing*, o Manual aconselha que se classifiquem os recursos de forma o mais específica possível dentro do esquema utilizado, mas mostrando-os nas classes mais gerais, o que permitirá quando o desenvolvimento do número de recursos o justificar uma visualização das respectivas sub-classes. A reorganização da hierarquia dos assuntos prevista na classificação justifica-se quando isso facilita a navegação do utilizador, nomeadamente pela subida de nível de alguns assuntos que se encontram representados em níveis mais baixos da hierarquia. Renomear os descritores associados às notações pode obviar as dificuldades para o utilizador final de uma terminologia demasiado específica e técnica. Em alguns casos pode verificar-se que o esquema de classificação seleccionado não é suficientemente detalhado em determinadas áreas ou omite categorias temáticas intimamente relacionadas com a cobertura temática seleccionada pelo serviço. Assim, pode ser necessário desenvolver o esquema estabelecido, prevendo-se as seguintes formas de desenvolvimento: adicionar uma subdivisão temática a algumas classes, sem alterar as classes existentes; adicionar facetas à classificação permitindo a subdivisão mais específica das classes; colar partes do sistema formando novas áreas para desenvolver a cobertura temática.

Enquanto que a função das classificações é a de agrupar recursos relacionados em categorias bem definidas, a utilização de palavras-chave destina-se a indexar de forma mais detalhada os conceitos presentes no documento, permitindo assim uma pesquisa mais refinada. Os autores reflectem neste particular sobre as vantagens e desvantagens da utilização de vocabulários controlados e não controlados, sugerindo a utilização de vocabulário controlado a ser suplementado por termos não controlados por forma a facilitar a pesquisa do utilizador que não conhece o termo controlado exacto. A característica hierárquica da organização vocabular de um tesouro é valorizada pelo Manual, dado que permite a navegação do geral para o particular, e ao mesmo tempo uma pesquisa por palavra-chave. Nas palavras dos autores “na prática, a combinação de pesquisa no Tesouros com a possibilidade de *browsing* uma pequena parte é a que, em

muitos casos, permite dar ao utilizador os melhores resultados.” (DESIRE INFORMATION GATEWAYS HANDBOOK)

As funcionalidades de *browsing* e pesquisa são abordadas como resultado dos métodos de classificação e indexação referidos, sugerindo-se a utilização de classificações para a criação de estruturas de *browsing*, bem como Tesouros com uma explícita e completa estrutura hierárquica. Indica-se que as funcionalidades de pesquisa por palavra podem ser realizadas sobre atributos formais dos documentos, como autores, títulos, aos quais se devem adicionar termos que especifiquem o assunto do recurso, bem como pesquisa em texto livre sobre as descrições. As classificações podem também ser utilizadas para este efeito, restringindo por exemplo a pesquisa a uma dada área temática.

Os atributos de tipo administrativo destinam-se a auxiliar a gestão da informação e são mais significativos para os gestores dos *Gateways* do que para o público. O manual DESIRE aconselha os seguintes: Manutenção do recurso, Data da adição do recurso, Data da última actualização do registo, Data da última alteração do recurso, Data da última revisão, Data de validade, Quem submete o recurso, Catalogador do recurso, Origem do registo (se o *Gateway* tem colaboradores), Detentor de direitos.

Assumindo que, por um lado, nem todos os elementos referenciados se adequam a todos os casos, e que, por outro, a sua utilização implica um esforço considerável quer na produção original do registo quer na manutenção, introduz-se a noção de “nível mínimo”, exemplificando com Dublin Core.

Atribuindo a responsabilidade de selecção dos atributos aos gestores do serviço e recusando portanto aconselhar um dado conjunto estandardizado de atributos, os autores definem, sob a forma de questões, um conjunto de critérios que permitirá guiar a selecção dos atributos mais adequados:

- Que campos de metadados serão utilizados vantajosamente pelos utilizadores?

Considerar os potenciais utilizadores, bem como a natureza dos recursos descritos

- Como vão os utilizadores pesquisar esses campos de metadados?

Esta variável afectará quer os atributos seleccionados quer as regras de catalogação a implementar.

- Que campos de metadados serão visualizados pelos utilizadores?

- Que campos de metadados serão utilizados para a gestão da informação e como?

- Que campos de metadados serão opcionais?

- Os metadados serão visualizados no mesmo formato em que são armazenados?

- Capacidades do software utilizado para gerar automaticamente atributos e quais?

- Quem vai gerar os campos de metadados e quais?

Os metadados podem ser suplementados por vários intervenientes, por exemplo:

- Fornecedores da informação
- Utilizadores
- Catalogadores
- Editores
- *Staff* do serviço
- Outro *Gateway* com o qual se colabora
- Geração automática

- Se utilizadores ou fornecedores da informação podem submeter recursos, que tipo de informação devem fornecer?

- Em que língua os metadados vão ser disponibilizados?

A selecção do formato de metadados adequado à estruturação / organização da informação descritiva depende de vários factores, sociais e tecnológicos, sendo que estes serviços adoptam tendencialmente formatos simples de carácter genérico minimamente estruturados, como Dublin Core. Este tipo de formatos apresenta essencialmente a vantagem da simplicidade, os seja, são da fácil criação e manutenção. A existência de algum nível de estruturação da informação facilita, por outro lado, a interoperabilidade e a conversão entre formatos. Sempre que isso seja considerado uma vantagem competitiva para o serviço, justificar-se-á a utilização de formatos mais complexos de metadados, com um nível de especialização mais exigente.

Num ambiente de rápida mutação, mas também devido à relativa novidade deste tipo de serviços, aconselha uma especial preocupação na garantia de condições de evolução. Estes serviços necessitam, assim, de garantir a flexibilidade necessária para lidar com novos elementos e novos formatos, bem como com a importação e exportação em vários formatos. Esta garantia pode ser conseguida se forem seguidas directrizes gerais para o conteúdo dos metadados. Estas directrizes devem incluir recomendações que assegurem:

- Concordância entre *Gateways* que cooperam relativamente a formatos de metadados e regras para o conteúdo
- *Gateways* devem implementar usos locais por meio de processamento local e não incorporá-lo nos dados
- Devem existir o mínimo de variantes locais em relação ao standard
- Colaboração com outros *Gateways* para que a migração possa tirar partido de economias de escala.

### **3.2.2 As Regras de Catalogação ROADS (Resource Organization and Discovery in Subject- Based Services)**

A construção de grande parte das linguagens da descrição dos serviços em análise tem por base, quer blocos normativos com origem no universo tradicional das bibliotecas, quer adaptações dessas regras a estas novas tipologias e ambientes documentais. No primeiro caso estão as *Anglo-American Cataloging Rules*, 2ª edição de 1988 com todas as alterações verificadas até 2001, particularmente relevante para este caso a alteração completa do capítulo 9 renomeado “*Electronic Resources*”.<sup>16</sup> No segundo caso, encontram-se um manual para a catalogação de recursos electrónicos elaborado por Nancy B. Olson no âmbito da OCLC, cuja primeira edição é de 1995<sup>17</sup> e as directrizes para a catalogação desenvolvidas por Rebeca Bradshaw para o *Gateway ADAM*<sup>18</sup>, provavelmente elaboradas em 1995. Integrando estas diversas construções bem como as ISBD (ER) (*International Standard Bibliographic Description. Electronic Resources*) estão as directrizes desenvolvidas no âmbito da aplicação ROADS, (com uma primeira versão em Janeiro de 1998) (DAY: 1999) que se revestem de uma particular importância neste contexto na medida em que muitos destes serviços utilizam o pacote

---

<sup>16</sup> Cf. Joint Steering Committee for Revision of Anglo-American Cataloging Rules – *A brief history of AACR*. Disponível em: <http://www.nlc-bnc.ca/jsc/history.html>

<sup>17</sup> Olson, Nancy B. – *Cataloguing Internet resources: a manual and a practical guide*. Disponível em: <http://www.oclc.org/oclc/man/9256cat/toc.htm>

<sup>18</sup> Bradshaw, Rebeca – *Cataloguing rules for the ADAM database: a procedural manual*. Disponível em: <http://www.adam.ac.uk/adam/reports/cat>.

aplicacional ROADS para a construção do catálogo. ROADS resulta de um projecto eLib, tendo apresentado os primeiros resultados em 1997, e é constituído por um conjunto de software e regras standardizadas “desenhado para ajudar a estabelecer e manter *Gateways* de informação para todos os tipos de recursos Internet, incluindo sítios WWW, serviços baseados em Telnet, sítios FTP e listas de discussão.” (A Guide to ROADS).

Todos os serviços, excepto HUMBUL, se baseiam de alguma forma nas Regras de Catalogação definidas para ROADS. SOSIG, BIOME, PSIGATE porque utilizam o software aplicacional ROADS e EEVL por ter construído um normativo próprio mas mapeado com ROADS.

As regras de catalogação desenvolvidas no âmbito do projecto ROADS destinam-se não só a normalizar a utilização das aplicações criadas, mas também a possibilitar a pesquisa cruzada entre os vários serviços que utilizam esta aplicação. Por isso, elas são mais do que simples regras de preenchimento de campos de informação de uma base de dados, traduzem a construção de uma verdadeira linguagem de metadados, definindo os atributos da informação a descrever, um dado vocabulário, uma semântica e uma sintaxe.

No que diz respeito aos metadados descritivos, estamos perante uma linguagem onde predominam os atributos relativos à **Obra** (Cf. Tabela 2), ou seja, ao conteúdo intelectual do recurso, sem deixar de incluir um significativo número de atributos relativamente ao **Documento**, ou seja, ao invólucro da informação.

**Tabela 2 – Atributos relativos à Obra em ROADS**

<b>Atributos relativos à "Obra"</b>	<b>Definição</b>
Título alternativo	Um título alternativo ao título ou ao título abreviado
Autor	Ver entidade Autor
Citação	Citação preferida para o recurso quando utilizado por outras Obras
Comentários	Comentários do catalogador
Descrição	Descrição ou resumo
Discussão	Fora de discussão apropriados para o recurso
Palavras-Chave (Nomes)	Palavras-chave para nomes de pessoas
Palavras-Chave (Organização)	Palavras-chave para nomes de organizações
Palavras-Chave (Assunto)	Palavras-chave de assunto
Língua	Língua do conteúdo intelectual
Título abreviado	Uma forma abreviada do título, por exemplo, acrónimo
Fonte	Informação sobre a versão definitiva do recurso
Linguagem do descritor de Assunto	Designação da linguagem utilizada
Descritor de Assunto	Termos da linguagem utilizada
Título	Título completo do recurso
Versão	Designativo da versão do recurso
Categoria	Tipo de recurso

Fonte: ROADS Cataloging Guidelines

**Tabela 3 – Atributos relativos ao Documento em ROADS**

<b>Atributos relativos ao “Documento”</b>	<b>Descrição</b>
Política de Acesso	Políticas de acesso ao serviço
Tempos de Acesso	Períodos de tempo preferidos ou mandatórios para o acesso
Editor	Ver entidade Organização
Proprietário do recurso	Ver entidade Organização
Patrocinador	Ver entidade Organização
Administrador do recurso	
Autenticação	Detalhes de autenticação como <i>logins</i> e <i>palavras-chave</i>
Conjunto de caracteres	Conjunto de caracteres utilizado
Política de “preço”	Política de preços para acesso a um recurso (se apropriado)
Direitos de autor	Declaração de direitos de autor
Data de Criação	Data de criação do recurso
Formato	Formato em que o recurso está disponível
ISBN / ISSN	Números normalizados
Estado de publicação	(sem descrição)
Registo	Detalhes de registo para acesso ao recurso se este não for disponibilizado livremente
Requisitos	Descrição de qualquer requisito de hardware ou software necessário à utilização do recurso
Tamanho	Extensão do recurso em bytes
URI	URI/URL

Fonte: ROADS Cataloging Guidelines

Os elementos relativos à **Obra** organizam-se à volta de três grandes tipos de atributos da informação: **autores**, **títulos** e **assuntos**. Os elementos relativos ao **Documento**, por sua vez, revelam a importância atribuída a três grandes domínios: localização e acessibilidade, publicação.

Autores pessoas físicas e organizações responsáveis pelo conteúdo, bem como tipos diferentes de responsáveis pela edição, disponibilização e manutenção do recurso são considerados, para além de atributos das entidades **Obra** e **Documento**, entidades em si mesmas (Cf. Tabela 3). Para a descrição das entidades autores pessoas físicas ou organizações são prescritos vários elementos que visam uma caracterização relativamente exhaustiva da entidade, o que traduz, por um lado, a tentativa de ultrapassar as dificuldades de identificação de autorias neste ambiente, mas também a possibilidade de estabelecer pontes de ligação entre autores e “leitores”, e por outro, a necessidade de distinguir entidades semelhantes através da junção de informação complementar ao nome que permite resolver eventuais ambiguidades. Quer num quer noutro caso, o conceito de responsabilidade é alargado a funções relacionadas com aspectos específicos da produção da informação: incluem-se neste conceito funções como Editor, Administrador do Sítio, Proprietário do Sítio, Patrocinador do Sítio, *Leaders* e parceiros de projectos. Os atributos da informação a descrever organizam-se à volta de três grandes áreas: identificação onomástica (dada pelo elemento nome), identificação profissional (dada pelo seguintes elementos: Departamento onde o indivíduo trabalha, Título profissional, Fax, Endereço Postal, Telefone) e identificação pessoal (dada pelos seguintes elementos: Endereço de correio electrónico, Fax, Telefone, Endereço postal). No caso das organizações, excluem-se evidentemente os elementos relativos à identificação pessoal, mas acrescentam-se os relativos a Cidade e Estado.

O vocabulário prescrito para o conjunto dos elementos é sobretudo vocabulário derivado que deve ser seleccionado a partir da fonte prescrita da informação. A noção de fonte prescrita da informação é concretizada através da identificação dos locais possíveis para informação, com especial referência à identificação do título. As fontes prescritas de informação podem oscilar, por ordem de importância, desde o ecrã inicial ou uma visualização similar, até à atribuição de um título pelo catalogador.

O vocabulário atribuído divide-se em duas categorias no que respeita ao controle terminológico: vocabulário em texto livre ou vocabulário controlado. O número de elementos descritivos com vocabulário controlado é bastante reduzido: só nos casos das datas (mesmo assim com a possibilidade de seleccionar de entre vários formatos possíveis todos eles válidos), da língua, do tamanho e versão é sugerido uma determinada forma de controlo vocabular. É no caso dos metadados de tipo administrativo que existe um maior controlo vocabular, visto que grande parte do vocabulário é suplementado de forma automática. O controlo vocabular para os assuntos é dado pela utilização da linguagem documental, embora só seja claramente estipulada a utilização de uma linguagem deste tipo no caso do elemento Descritor.

Ainda no que diz respeito ao vocabulário, verifica-se uma completa ausência de regras de forma, por exemplo, quanto à utilização de abreviaturas, maiúsculas ou formas linguísticas dos nomes, exceptuando o caso dos títulos em que se prescreve a capitalização dos nomes próprios.

Do ponto de vista da semântica são praticamente inexistentes quaisquer regras, exceptuando o caso da responsabilidade intelectual sobre o recurso, na qual o número de elementos de informação sobre um autor permite claramente distinguir entidades semelhantes.

As regras sintácticas são muito reduzidas e acontecem apenas nos seguintes casos: separação de palavras-chave por vírgula, separação entre título e subtítulo pela utilização de dois pontos, forma invertida do nome do autor com separação entre os dois elementos por vírgula.

Mais recentemente foi desenvolvido um *template* específico para a descrição de recursos Internet, designado por *RESOURCE*<sup>19</sup>, particularmente pensado para o contexto RDN e para ser conforme com Dublin Core. Neste *template* são definidos atributos para a entidade **Obra, Documento, Autor, Editor; Assunto** bem como elementos relativos à administração da descrição.

**Tabela 4 – Atributos relativos à Obra no *template* RESOURCE**

<b>Atributo / Etiqueta</b>	<b>Definição</b>
Título	O nome dado a um recurso
Título alternativo	Qualquer forma de nome dado a um recurso que é usado como substituto ou alternativo ao título formal
Palavras-chave	Palavras ou frases-chave que descrevem o conteúdo do recurso
Descrição	Uma descrição do conteúdo do recurso. A

<sup>19</sup> Disponível em <http://www.ukoln.ac.uk/metadata/roads/templates/resource.html>

	descrição pode incluir, mas não está limitada a: resumo, índices ou descrição em texto livre do recurso
Categoria	Natureza ou género do recurso. Tipos incluem termos que descrevam categorias gerais, funções, géneros ou níveis de agregação de conteúdos
Fonte	Referencia ao recurso do qual o actual deriva
Língua	Língua do conteúdo intelectual do recurso
Relação	Referência a um recurso relacionado
Cobertura	Extensão ou âmbito do conteúdo do recurso. Cobertura incluirá tipicamente localização espacial (nome de lugar ou coordenadas geográficas), período de tempo (nome de um período, data ou intervalo de datas) ou jurisdição (por exemplo, o nome de uma entidade administrativa)
Comentários	

Fonte: *Template Resource*

Ao conjunto de atributos definidos para a **Obra** (Tabela 4) devem acrescentar os atributos das entidades responsabilidade intelectual e assunto, já que são considerados duplamente, quer enquanto entidades individuais, quer como atributos da entidade **Obra**.

Quando comparados estes atributos com os que são definidos pelas Regras de Catalogação (Cf. Tabela 2) verifica-se quer o acrescento de novos elementos, quer a ausência de outros, quer ainda alguns níveis de alteração de um ou outro elemento.

No primeiro caso encontram-se os atributos **relação** e **cobertura**. No segundo caso, desaparecem os elementos: **citação**, **discussão**, **título abreviado**, **versão**, **proprietário do recurso** e **patrocinador**. No último caso encontra-se o atributo **palavras-chave** que perde os qualificadores nomes individuais e organizações.

No que diz respeito ao **Documento** (Cf. Tabela 5), verifica-se uma redução drástica do conjunto de atributos definidos anteriormente, a saber: **política de acesso**, **tempos de acesso**, **autenticação**, **conjunto de caracteres** utilizados, **política de preço**, **estado de publicação**, **registo**, **requisitos**, **ISBN/ISSN**, **tamanho**, **detentor do recurso**, **patrocinador do recurso**. A queda destes atributos deve ser compreendida tendo em conta vários factores, para além da adaptação a Dublin Core: em primeiro lugar, quando as regras de catalogação ROADS foram elaboradas os serviços baseados em Telnet ocupavam uma importância maior do que os que se baseavam na WWW, por isso, a importância de considerar atributos como Política de acesso, Tempos de Acesso, Autenticação

**Tabela 5 – Atributos relativos ao Documento no *template RESOURCE***

<b>Atributo / Etiqueta</b>	<b>Descrição</b>
Data - Criação	A data em que o recurso foi criado
Formato	Manifestação física ou digital do recurso. Tipicamente, formato pode incluir o tipo de media ou as dimensões do recurso. Formato pode ser utilizado para determinar os requisitos de software, hardware ou outro equipamento necessário para visualizar ou operar com o recurso.

URI	Identificador Uniforme para o recurso
Direitos de autor	Informação sobre os direitos detidos sobre o recurso

Fonte: *Template Resource*

No que diz respeito às entidades ligada ao conteúdo intelectual verifica-se também uma drástica diminuição dos atributos, ficando reduzidos à identificação onomástica e ao endereço electrónico (Cf. Tabelas 6).

**Tabela 6 – Atributos relativos a responsabilidade intelectual sobre o recurso no *template* RESOURCE**

<b>Atributos / Etiquetas</b>	<b>Descrição</b>
Autor – Nome	O nome da pessoa ou organização responsável, em primeiro lugar, pela produção do conteúdo do recurso
Autor – endereço electrónico	O endereço electrónico da pessoa ou organização responsável, em primeiro lugar, pela produção do conteúdo do recurso
Contribuinte – Nome	O nome da pessoa ou organização responsável por contributos para o conteúdo do recurso
Contribuinte – endereço electrónico	O endereço electrónico da pessoa ou organização responsável por contributos para o conteúdo do recurso
Admin. – Nome	O nome da pessoa ou organização responsável pela manutenção do recurso na sua actual localização
Admin. – endereço electrónico	O endereço electrónico da pessoa ou organização responsável pela manutenção do recurso na sua actual localização

Fonte: *Template Resource*

As entidades **Editor** e **Administrador do Recurso** (Cf. Tabela 7, Tabela 8), se bem que estivessem previstas nas Regras de Catalogação ROADS, são aqui autonomizadas como categorias individuais. Se comparados os atributos que possuíam na entidade onde anteriormente estavam incluídos, ou seja, a entidade **ORGANIZAÇÃO**, verifica-se também uma drástica diminuição dos seus atributos.

**Tabela 7 – Atributos relativos ao editor no *template* RESOURCE**

<b>Atributos / Etiquetas</b>	<b>Definição</b>
Editor – Nome	O nome da pessoa ou organização responsável por tornar o recurso disponível
Editor – endereço electrónico	O endereço electrónico da pessoa ou organização responsável por tornar o recurso disponível

Fonte: *Template Resource*

**Tabela 8 – Atributos relativos ao administrador do recurso no *template* RESOURCE**

<b>Atributos / Etiquetas</b>	<b>Definição</b>
Admin - Nome	O nome da pessoa ou organização responsável pela manutenção do recurso na sua localização actual.
Admin – endereço electrónico	O endereço electrónico da pessoa ou

	organização responsável pela manutenção do recurso na sua actual localização
--	--

Fonte: *Template Resource*

A consideração de uma entidade assunto enquanto objecto descritivo apresenta-se como uma novidade em relação às Regras de Catalogação ROADS e reúne os dois atributos mínimos (Tabela 9) que permitem identificar o conteúdo atribuído. Note-se que não é considerada a possibilidade de utilização de termos directamente retirados da linguagem natural, sem estarem enquadrados numa linguagem controlada.

**Tabela 9 – Atributos relativos ao assunto no *template RESOURCE***

<b>Atributos / Etiquetas</b>	<b>Definição</b>
Assunto – Descritor	Um termo de um vocabulário controlado ou de um esquema de classificação formal que descreve o conteúdo do recurso
Assunto – Descritor - Esquema	

Fonte: *Template Resource*

Em relação à forma como este tipo de entidades era vista nas Regras de Catalogação ROADS, verifica-se aqui uma deslocação da perspectiva no sentido de realçar a função desempenhada pelos indivíduos/organizações na produção e manutenção do recurso e não tanto o tipo de entidade, individual ou colectiva.

As especificações do *template* em análise não identificam qualquer definição relativamente a três dos parâmetros que temos vindo a analisar e que caracterizam qualquer linguagem de metadados, ou seja, vocabulário, semântica e sintaxe. Na medida em que pertence ao universo ROADS, podemos concluir que se aplicam as mesmas regras já analisadas.

O aparecimento do *template RESOURCE* representa uma clara diminuição da especificidade descritiva prevista pelas Regras de Catalogação ROADS numa expressa aproximação à simplicidade de Dublin Core na sua versão simplificada. A análise das linguagens produzidas por cada serviço, nomeadamente aqueles que explicitamente se baseiam em ROADS ou por utilizarem o pacote aplicacional ou por terem realizado uma aproximação normativa, permitirá avaliar o nível de influência destas propostas.

Se bem que tenha surgido já anteriormente, a relação entre ROADS e DESIRE é próxima, sendo este pacote aplicacional apresentado como uma forma de concretizar a perspectiva desenvolvida no manual para os *Subject Gateways* (Sítio *Web DESIRE*: 1999).

Se comparadas estas regras com os códigos catalográficos tradicionais verifica-se a existência de uma grande número de similitudes, quer no que diz respeito aos atributos a incluir na descrição, quer à utilização de vocabulários controlados para a representação do conteúdo (CHAPMAN: 1998). As principais diferenças dizem respeito à descrição dos documentos enquanto entidades físicas, a um menor nível de complexidade e especificidade das regras de descrição e a uma adaptação da semântica e conteúdo de determinados atributos clássicos às realidades dos recursos em rede.

### 3.3. Origem, evolução e características actuais dos Gateways

#### 3.3.1. HUMBUL: humanities hub<sup>20</sup>

HUMBUL é o acrónimo de *HUMANities BULLETin*, o que é ainda hoje revelador da sua origem, a meio dos anos 80, como *Bulletin Board* destinado à comunidade universitária das Ciências Humanas do Reino Unido. Naquela fase, o serviço criado na Universidade de Bath e da responsabilidade do *Office for Humanities*, pretendia sobretudo ir ao encontro dos interessados pelas potencialidades da aplicação das tecnologias da informação e da comunicação às humanidades.

Em 1991 o serviço é transferido para a Universidade de Oxford, ficando a cargo do *Oxford University Computing Service* (OCUS). Esta transferência institucional não produz quaisquer alterações de conteúdo ou de missão.

Em 1994, o serviço evolui para um *Gateway*, implementando uma solução baseada num conjunto de páginas estáticas em HTML. O crescimento consecutivo do número de recursos descritos torna esta solução técnica rapidamente inadequada, quer pelas dificuldades de acesso à informação sentidas pelos utilizadores, quer pelas dificuldades de manutenção.

Em 1997, o serviço evolui então para uma base de dados, permitindo um rápido crescimento, uma melhoria da qualidade de acesso, sobretudo pela possibilidade de implementar a pesquisa por palavra-chave, mas também uma gestão mais flexível e eficaz.

A partir desse momento, a evolução tem vindo a verificar-se sobretudo pelo contínuo crescimento de recursos descritos e pela inserção de novos serviços com o objectivo de, por um lado, aumentar e diversificar as formas de aceder aos conteúdos, e por outro, desenvolver as facilidades de customização individual do acesso e dos resultados. Nestes domínios, as grandes alterações<sup>21</sup> prendem-se com a adesão à RDN, a implementação de um canal de RSS (*Rich Site Summary*) e com a criação de funcionalidades ligadas à personalização dos serviços. Neste último caso, tal é concretizado através da criação, em 2001, do serviço “*My Humbul*” (METCALFE: 2001). Por via deste novo serviço, os utilizadores podem, através da gravação das suas pesquisas ou pela selecção de assuntos a partir da árvore de assuntos do serviço, serem alertados periodicamente de novos recursos descritos na base de dados. É também possível a personalização do interface de pesquisa. A implementação do canal de RSS que permite transferir as descrições de sítios recentemente adicionados ao serviço para qualquer sítio *Web*.

Outras alterações traduziram-se na disponibilização de informação sobre ofertas de emprego na área das humanidades, através do serviço nacional de emprego online; e na transparência relativamente aos conteúdos das descrições, através da identificação do seu autor.

---

<sup>20</sup> Disponível em <http://www.humbul.ac.uk>

<sup>21</sup> Para um lista completa das alterações realizadas por este serviços nos anos de 2001-2002 ver “Recente changes”. Disponível em <http://www.humbul.ac.uk/about/changes.html>

Pensado para ir ao encontro das necessidades de informação da comunidade de investigadores e estudantes do Reino Unido na área das Humanidades, HUMBUL é actualmente da responsabilidade da Universidade de Oxford, através do OCUS e dos *Oxford Libray Services*, em estreita colaboração com o *Arts and Humanities Data Service* (AHDS). Para além do conjunto de profissionais responsáveis pela manutenção do serviço<sup>22</sup>, HUMBUL conta ainda com a colaboração de um importante conjunto de instituições do meio universitário<sup>23</sup> que contribuem para o crescimento do catálogo de recursos. Para além deste conjunto de cooperantes, o serviço dispõe de um Comité de Aconselhamento<sup>24</sup>, cujos objectivos são: fornecer aconselhamento especializado no que concerne à manutenção e desenvolvimento do serviço; monitorizar as actividades do serviço; contribuir para a definição de estratégias de desenvolvimento.

O catálogo de recursos electrónicos disponíveis via Internet, a espinha dorsal do serviço, conta actualmente com 4.868 registos<sup>25</sup>, seleccionados de acordo com uma clara política de desenvolvimento e manutenção da colecção<sup>26</sup> que estabelece que o âmbito temático dos recursos a incluir passa pelas seguintes áreas: História, Literatura e Língua Inglesa, Línguas e Literaturas Modernas, Linguística, Arqueologia, História Antiga e Línguas e Literaturas Clássicas, Religião e Teologia e, finalmente, Filosofia. Neste âmbito, é também definida uma tipologia de recursos a incluir, bem como critérios de selecção e de gestão / manutenção das colecção, neste último caso com regras e procedimentos quanto à manutenção da descrição de um recurso na base de dados e/ou respectiva anulação. Os recursos são descritos de acordo com um conjunto bem definido de regras de descrição, corporizadas num Manual de Catalogação (FRASER:2001).

O utilizador pode aceder ao conteúdo por duas vias: pesquisa por palavra-chave e navegação numa árvore de assuntos através de um sistema de hiperligações. A pesquisa por palavra-chave apresenta funcionalidades reduzidas: não é possível determinar o campo objecto de pesquisa (por exemplo, autor, descrição, assunto, etc.), os operadores de pesquisa apresentam-se apenas sob as formas de “Any” (qualquer das palavras), “All” (todas as palavras) e “Phrase” (frase exacta). Recentemente, está disponível em cada um dos assuntos a possibilidade de restringir a pesquisa a esse assunto ou pesquisar todo o catálogo<sup>27</sup>. A aposta parece, assim, centrar-se mais nas possibilidades de *browsing* dos recursos proporcionadas pela árvore de assuntos já referida.<sup>28</sup>

---

<sup>22</sup> Para uma lista completa dos elementos da equipa HUMBUL ver Humbul Team. Disponível em <http://www.humbul.ac.uk/about/staff.html>

<sup>23</sup> A lista completa das instituições pode ser consultada em Introduction to HUMBUL. Disponível em <http://www.humbul.ac.uk/about/index.html>

<sup>24</sup> Para mais informação sobre a constituição do Comité de Aconselhamento ver HUMBUL advisory committee. Disponível em <http://www.humbul.ac.uk/about/advisory.html>

<sup>25</sup> Dados retirados em 03/02/27

<sup>26</sup> HUMBUL – Collection development policy. Disponível em <http://www.humbul.ac.uk/about/colldev.html>

<sup>27</sup> O anúncio desta funcionalidade data de 03/02/21. Cf. Subject search. Disponível em: <http://www.humbul.ac.uk/about/changes.html>

<sup>28</sup> Dado que a organização dos assuntos para efeitos de *browsing* dos recursos será objecto específico de desenvolvimento em capítulo posterior, aqui é apenas feita uma breve referência.

### 3.3.2. SOSIG: social science information gateway<sup>29</sup>

O *Social Science Information Gateway* nasce em 1994 no *Institute for Learning and Research Technology* (ILRT) da Universidade de Bristol, com o objectivo de apoiar investigadores, estudantes e bibliotecários na área das ciências sociais na localização e identificação de recursos electrónicos de qualidade que pudessem apoiar o ensino e a investigação. Nasce, desde logo, como catálogo de recursos electrónicos disponíveis na Internet, baseado em descrições que os utilizadores podem pesquisar ou navegar através de uma árvore de assuntos, implementando um modelo que será seguido posteriormente por muitos outros.

A partir de 1996, SOSIG participa activamente no projecto DESIRE, o que constitui indubitavelmente o ponto de partida para o grande conjunto de transformações que se têm verificado a partir daí e que tornou este serviço, de certa forma, no agente da concretização do “paradigma” definido por esse projecto comunitário.

Logo em 1997, o serviço conhece várias alterações na configuração do interface de pesquisa e na introdução do tesaurus HASSET (*Humanities And Social Science Electronic Thesaurus*), o qual os utilizadores poderão pesquisar como uma das formas de aceder aos recursos (HOOPER: 1997).

Em 1998 é lançado um novo serviço para além do catálogo, designado por *Social Science Grapevine* e destinado a dar corpo à ideia de unir a comunidade de práticas dos cientistas sociais. Através deste serviço, o utilizador tem acesso a informação sobre eventos, cursos, departamentos universitários na área das ciências sociais, pode publicar o seu currículo, aceder a informação sobre indivíduos que trabalham na mesma área (HUXLEY: 1998).

Mas é sem dúvida o ano de 1999 que trará o mais importante conjunto de transformações, embora os reflexos para o público se façam sentir só a partir do início de 2000 (KITCHEN: 2000). Para além de um novo *design*, destacam-se, em termos de interface público, a pesquisa por palavra-chave sempre disponível em qualquer página do serviço, um mapa completo com todas as secções e a possibilidade de pesquisar o motor de pesquisa das ciências sociais.<sup>30</sup> O alargamento e aprofundamento das temáticas cobertas pelo serviço conhece também um novo desenvolvimento através da cooperação com outro *Gateway* da mesma universidade, especializado em economia e negócios (*Biz/ed*) e pelo desenvolvimento da área do Direito, designado agora por *Law Gateway*, resultado da cooperação com o *Institute of Advanced Legal Studies* (IALS) e com a *Bristol Law Library*. A partir daqui SOSIG significará *The Social Science, Bussiness and Law Hub*. Nesta fase é também criada, através do serviço *My account*, a possibilidade de configurar de forma personalizada o acesso à informação. Através da definição de um perfil de interesses que pode ser conjugado com entradas na *Grapevine*, o utilizador pode: receber semanalmente notificação, por correio electrónico, de novos recursos descritos, conferências, cursos, etc. encontrar profissionais na mesma área; aceder a conferências e cursos relacionados com os seus interesses; configurar uma

---

<sup>29</sup> Disponível em <http://www.sosig.ac.uk>

<sup>30</sup> Literalmente “Social Science Search Engine” é uma base de dados autónoma relativamente ao catálogo e construída a partir da pesquisa de hiperligações referenciadas nos sítios descritos no catálogo.

página *Web* para aceder a canais de notícias e a actualizações a partir de serviço seleccionados.

Finalmente em 2001, assiste-se ao lançamento de uma secção sobre Estudos Europeus, que tem como tema central a Europa enquanto região (WOODCOCK: 2001).

Para além dos serviços que já foram referenciados (*Grapevine* e *My account*) SOSIG conta actualmente com um catálogo de 23.326<sup>31</sup> registos e com um motor de pesquisa com 50.000 páginas *Web*. Os recursos descritos no catálogo são seleccionados de acordo com uma política que define o âmbito dos assuntos a incluir, as fontes de informação, os níveis de dificuldade, o tipo de recursos, aspectos linguísticos, geográficos e de acessibilidade.<sup>32</sup> Após serem identificados, os recursos são avaliados de acordo com critérios de selecção estabelecidos<sup>33</sup> e geridos pela aplicação de procedimentos estabelecidos numa política de gestão da colecção.<sup>34</sup> As áreas temáticas actualmente cobertas por este serviço são: Negócios (incluindo a Gestão), Economia, Educação, Ciências do Ambiente, Etnografia, Etnologia, Antropologia, Geografia, Administração Pública e Governo, Direito, Filosofia, Política, Psicologia, Ciências Sociais em geral, Serviço Social, Sociologia, Estatística e Estudos sobre as Mulheres. Após o processo de selecção, os recursos são descritos de acordo com regras bem definidas, corporizadas através de um Manual de Catalogação (HIOM: 1999).

A informação sobre os recursos descritos no catálogo pode ser acedida por quatro vias: a pesquisa por palavra-chave, a navegação na árvore de assuntos<sup>35</sup>, a pesquisa no tesaurus e o acesso ao motor de pesquisa das ciências sociais.

A pesquisa por palavra-chave apresenta duas hipóteses: a pesquisa simples e a pesquisa avançada. A pesquisa simples assume por defeito o operador E (AND), mas os operadores OU e NÃO podem também ser utilizados, bem como a truncatura e a pesquisa por frase exacta. A pesquisa é executada sobre qualquer um dos campos de informação que compõem a descrição bibliográfica do recurso. A pesquisa avançada, para além das possibilidades já referidas na modalidade anterior, permite ainda: restringir a uma das partes do catálogo, ou seja, às grandes áreas de assunto do directório, restringir o campo de informação sobre o qual a pesquisa vai actuar, restringir ao tipo de recurso que se pretenda, optar por fazer actuar a truncatura, a ordenação dos registos de acordo com a sua relevância para a pesquisa (contada pelo número de vezes que o termo pesquisado aparece no registo), determinar a possibilidade de erros de ortografia (num máximo de 2), configurar a visualização do resultados em termos de nº de resultados mostrados em cada página, e a possibilidade de ver apenas os títulos ou estes e as respectivas descrições.

A pesquisa do tesaurus pode ser seleccionada directamente pelo utilizador ou ser-lhe sugerida sempre que o resultado de uma pesquisa não apresenta resultados. O tesaurus pode ser seleccionado na sua globalidade ou fazer actuar uma restrição às duas

---

<sup>31</sup> Dados retirados em Fevereiro de 2003

<sup>32</sup> SOSIG – Scope policy. Disponível em <http://www.sosig.ac.uk/desire/escope.html>

<sup>33</sup> SOSIG – Selection criteria. Disponível em: <http://www.sosig.ac.uk/desire/ecrit.html>

<sup>34</sup> SOSIG – Collection management policy. Disponível em:  
<http://www.sosig.ac.uk/desire/collect.html>

<sup>35</sup> Dado que este assunto será objecto de análise pormenorizada noutra capítulo, aqui á apenas feita esta referência.

seguintes áreas: Governação, Política e Antropologia ou Serviço Social. É ainda possível seleccionar a truncatura ou um dos seguintes operadores: E e OU.

A utilização do motor de pesquisa das ciências sociais permite restringir a busca aos campos de título e descrição, e as mesmas opções relativamente a truncatura, ordenação, e visualização já referenciadas na pesquisa avançada.

A visualização por defeito dos resultados mostra os seguintes campos de informação: título, descrição, palavras-chave e endereço. Através da opção registo completo (*full record*) é possível aceder a mais informação: secções onde o registo foi classificado, tipo de recurso e língua. As secções onde o registo foi classificado traduzem-se em hiperligações para a referida secção na estrutura do directório, permitindo assim ao utilizador obter informação sobre outros registos disponíveis para o mesmo assunto.

### **3.3.3. PSIGATE: the physical sciences information gateway<sup>36</sup>**

PSIGATE, *the physical sciences information gateway* foi apresentado ao público em Setembro de 2001 para cumprir a missão de disponibilizar o acesso a recursos electrónicos de qualidade acessíveis via Internet para o público alvo de estudantes, investigadores e profissionais das ciências físicas, especialmente nas seguintes áreas: Astronomia, Química, Ciências da Terra, Física, Ciências dos Materiais e História e Política científicas.

O serviço é da responsabilidade do CALIM, consórcio das bibliotecas universitárias de Manchester, que inclui: *John Rylands University Library of Manchester, Manchester Metropolitan University Library, UMIST Library, University of Salford Library* e a *Manchester Business School Library*.

Os recursos *Web* a incluir no catálogo, que conta actualmente com 7.100<sup>37</sup>, são seleccionados de acordo com uma clara política que define o âmbito temático dos recursos a incluir em cada temática geral, o tipo de recursos a incluir e excluir, as fontes de informação para obter esses recursos, a audiência que deve ser tida em conta, a cobertura geográfica e a língua em que o recurso está escrito<sup>38</sup>. Uma vez seleccionados, os recursos são avaliados de acordo com critérios bem definidos que abrangem as áreas do contexto de produção do recurso, do seu conteúdo e, finalmente do seu formato.<sup>39</sup> Após concluída a avaliação, os recursos são descritos na observância de um conjunto bem definido de regras de descrição, estruturadas num Manual de Catalogação.<sup>40</sup>

O acesso aos recursos pode ser realizado através da pesquisa por palavra-chave ou através do *browsing* dos recursos a partir de uma estrutura hierárquica, subdividida nas seis grandes áreas temáticas que compõem o serviço.

O acesso por palavra-chave apresenta três opções: uma pesquisa simples, uma pesquisa avançada, e uma pesquisa sobre um conjunto de bases de dados específicas. Nos dois

---

<sup>36</sup> Disponível em <http://www.psigate.ac.uk>

<sup>37</sup> Dados retirados em 03/02/27

<sup>38</sup> Cf. Resource scope: full version. Disponível em: <http://www.psigate.ac.uk/contentza.doc>

<sup>39</sup> Cf. Resource evaluation criteria. Disponível em: <http://www.psigate.ac.uk/content3.doc>

<sup>40</sup> Cf. Cataloguing Rules. Disponível em: <http://www.psigate.ac.uk/content4.doc>

primeiros casos a pesquisa actua sobre os campos título, palavras-chave e descrição. A pesquisa simples assume o operador E por defeito e executada sobre todos os assuntos. Na pesquisa avançada, é possível restringir de acordo com dois critérios: o assunto, podendo seleccionar-se um dos assuntos principais; e o tipo de recursos, através do acesso a uma extensa lista de tipos de recursos presentes no catálogo.

Em termos de apresentação e visualização dos resultados, estes são apresentados por ordem alfabética, estando disponíveis dois formatos de visualização: um, por defeito, contém o título do recurso, a descrição e o URL, e o outro, acedido a partir do anterior, disponibiliza ainda informação relativa a palavras-chave, assuntos, tipo de recurso, data da última modificação, e o detentor do copyright.

A última opção de pesquisa permite aceder a um conjunto de 7 bases de dados e instrumentos de pesquisa, quer mais generalistas quer mais específicos. O primeiro recurso disponível é uma base de dados que armazena o texto completo das páginas referidas pelos recursos descritos no catálogo. A pesquisa pode ser feita através de duas ou mais palavras-chave que serão procuradas, de acordo com a escolha do utilizador, no título, no texto completo ou em ambos. O segundo recurso disponível é constituído por um conjunto de bases de dados de patentes, informação sobre compostos químicos, espectroscopia, etc. O terceiro recurso permite a pesquisa cruzada de PsiGate, ZETOC (*British Library's Electronic Table of Contents of current journals and conference proceedings*), COPAC (um catálogo colectivo que reúne os catálogos de várias universidades do Reino Unido e da Irlanda e o da *British Library*) e finalmente a RDN, ou seja, todos os serviços que compõem esta rede através da funcionalidade *Resource Finder*. O resto dos recursos disponíveis tem origem no IoP (*Institute of Physics*) e na RSC (*Royal Society of Chemistry*).

Para além dos recursos referenciados, PsiGate disponibiliza ainda o acesso a um vasto conjunto de serviços específicos, a saber: Notícias, Bases de Dados, Lista de Recursos relacionados organizada de acordo com as temáticas globais do serviço, Lista de materiais de referência para cada um dos seis grandes temas do serviço, Lista de recursos de natureza educacional, Informação sobre emprego e carreiras.

### **3.3.4. EEVL: the Internet guide to engineering, mathematics and computing<sup>41</sup>**

EEVL, *the Internet guide to engineering, mathematics and computing* apareceu ao público, pela primeira vez em 1996 como um *Gateway* dedicado à engenharia. Desde essa data e até 1996 o serviço tem vindo a evoluir não só pelo crescimento do número de recursos descritos no catálogo, mas também pela adição de novos acessos a outros recursos pretendendo assim expandir o universo de possibilidades. De entre estas novas possibilidades destacam-se o motor de pesquisa para recursos em engenharia (*UK engineering search engine*) construído a partir dos recursos seleccionados para o catálogo indexando todo o conteúdo desses recursos, o motor de pesquisa para jornais electrónicos na mesma área (*Enginnering e-Journal search engine*), uma base de dados sobre a indústria intitulada *Recent advances in manufacturing* (RAM), e ao serviço *Offshore Engineering Information Service*, da responsabilidade da mesma Universidade.

---

<sup>41</sup> Disponível em <http://www.eevl.ac.uk>

Em 1999, o serviço incorpora recursos de outro *Gateway* sobre as ciências aeroespaciais – o AERADE, *gateway to quality aerospace and defence resource* – da responsabilidade da Universidade de Cranfield.

Em 2001 assiste-se à última grande transformação deste serviço, com a incorporação de recursos na área da Matemática e da Computação, assumindo então as suas características actuais. (MACLEOD: 2001)

Actualmente, EEVL é o resultado da parceria entre as Universidades de *Heriot-Watt*, responsável pela gestão global do serviço, *Birmingham*, *Cranfield* e *Ulster* e assume como missões principais a de disponibilizar o acesso a recursos de qualidade nas áreas da engenharia, matemática e computação para um público alvo maioritariamente de universitários e investigadores e a de ser o ponto focal nacional para o acesso a informação online nas referidas temáticas. A área das matemáticas é da responsabilidade do *Gateway MATHGATE*, que resulta da colaboração entre os serviços de informação da Universidade de *Birmingham* e do sítio *Maths Stats and OR Network*, um dos componentes da rede LTSN (*Learning and Teaching Support Network*). Os temas da computação são da responsabilidade da *Learning and Teaching Support Network for Information and Computer Services*, sediado na Universidade do *Ulster*. (MACLEOD: 2001 a)

Para além destas parcerias, o serviço conta ainda com um grupo de consultores para o desenvolvimento dos conteúdos que também contribui para o crescimento do catálogo, um grupo de aconselhamento geral (*Advisory Group*) e um grupo mais vocacionado para as questões relacionados com a infra-estrutura tecnológica do serviço (*Hub advisory group*).

Os recursos que incluem o catálogo de EEVL são seleccionados e avaliados de acordo com critérios definidos, que estabelecem: a abrangência temática em termos de assunto, tipos de recursos aceitável, nível de dificuldade, cobertura geográfica e língua, acesso e tecnologias envolvidas; a avaliação é ditada por um conjunto de critérios relativos ao contexto, conteúdo, usabilidade e características do sistema.<sup>42</sup>

Em termos de acesso à informação, EEVL possibilita aos utilizadores quatro possibilidades de pesquisa dos conteúdos através da utilização de palavras-chave: a pesquisa de recursos chave – trata-se de um parte do catálogo composta por recursos de maior qualidade nas várias áreas temáticas, no fundo, um filtro que permite restringir a pesquisa desde o início –, a pesquisa do catálogo, o acesso ao motor de pesquisa, e finalmente uma opção que conjuga as três primeiras hipóteses.

Como seria de esperar as opções de pesquisa para todo o catálogo e para os sítios chave são as mesmas: a pesquisa pode ser executada para todos os seguintes campos – título, URL, descrição e autor – ou apenas num deles; restrição pelas grandes áreas de assunto – engenharia, computação e matemática –; restrição por localização geográfica – Reino Unido ou Europa – ou alargamento a todo o Mundo; restrição por tipo de recursos disponibilizando uma extensa lista de tipologias de recursos sobre a qual podem ser utilizados os operadores booleanos E ou OU; possibilidade de utilizar os operadores booleanos básicos através das opções “Any”, “All”, “Phrase”; possibilidade de fazer

---

<sup>42</sup> Cf. EEVL collection development policy. Disponível em: <http://www.eevl.ac.uk/cdp.htm>

actuar a condição de pesquisar pela palavra exacta ou não, ou seja, possibilidade de utilização da truncatura; e finalmente a possibilidade de configurar a visualização dos resultados com uma única opção disponível que é a de visualizar apenas os títulos.

A pesquisa utilizando o motor (opção designada por *Sítios Web*) permite apenas a utilização dos operadores “Any” e “All”, bem como a truncatura e configuração da visualização já identificada. Em Novembro de 2002<sup>43</sup> é anunciada a disponibilização de três novos motores pesquisa, um por cada área de assunto, que podem ser acedidos a partir de cada uma das secções, ou seja, através da aplicação de um filtro temático sobre todo o índice é possível constituir índices individuais para cada uma das partes, já que, como vimos, o índice é constituído a partir do conteúdo dos recursos identificados para cada área temática. No caso do tema Engenharia, é disponibilizada ainda uma possibilidade de pesquisar os recursos da rede LTSN para este domínio.

Ao seleccionar a possibilidade de pesquisar ao mesmo tempo todas as vias descritas (opção designada por *Search all*), o utilizador pode utilizar os operadores “Any”, “All” e “Phrase”, bem como a truncatura e a opção de configurar a visualização.

Para além da pesquisa por palavras-chave, EEVL disponibiliza ainda a possibilidade de *browsing* dos recursos nas três grandes áreas temáticas.<sup>44</sup>

Para além da pesquisa no catálogo ou nos motores de pesquisa, EEVL apresenta ainda um extenso leque de serviços, a saber: uma área de notícias sobre as três temáticas base, as quais podem ser acedidas através de *browsing* das sub-categorias, o top dos 100 sítios mais acedidos em EEVL, a pesquisa da base de dados RAM, através de uma hiperligação que nos conduz ao sítio respectivo, a selecção de um conjunto de bases de dados bibliográficas relevantes, com hiperligação ao respectivo sítio; acesso a informação actualizada nas três grandes temáticas e nas seguintes áreas: recursos Internet, livros, artigos, seminários, congressos e comunicações; através de um conjunto variado de filtros sobre o catálogo estão disponíveis vários serviços temáticos: recursos sobre emprego, sobre educação no contexto da LTSN, sobre aprendizagem e ensino. Através da cooperação com o grupo informal de bibliotecários universitários da área das ciências e tecnologias, encontra-se disponível uma base de dados de contactos deste grupo profissional.

No final de 2002 (Dezembro)<sup>45</sup> é anunciada uma nova funcionalidade, a de, através de um canal de RSS, poder adicionar as últimas actualizações de EEVL a um outro sítio *Web*, depois de a partir de Outubro estar também disponível a possibilidade de implementar num qualquer sítio *Web* uma “caixa” de pesquisa para EEVL.

Em termos de visualização dos registos, o utilizador tem disponível, após a apresentação dos resultados, uma visualização por defeito que inclui título, URL e descrição, bem como o acesso ao registo completo que, para além da informação anterior, disponibiliza ainda: título alternativo (quando tal se aplica), palavras-chave, tipo de recurso, descritores da classificação, localização geográfica do recurso, país de origem do

---

<sup>43</sup> Notícia disponível em <http://www.eevl.ac.uk/pressrealease/pressrelenewWebsítios.htm>

<sup>44</sup> Dado que este assunto será objecto de desenvolvimento aprofundado, aqui faz-se apenas uma breve referência.

<sup>45</sup> Notícia disponível em <http://www.eevl.ac.uk/pressrealease/presslrss.htm>

recurso, língua, e ainda a possibilidade de comentar o registo ou aceder à sua codificação em XML.

### **3.3.5. BIOME: your guide to Internet resources in the health and life sciences<sup>46</sup>**

De entre todos os serviços que compõem a RDN, BIOME: *your guide to Internet resources in the health and life sciences* é o que de forma mais evidente concretiza o conceito de Hub, já que não tem conteúdo específico em si próprio, mas resulta completamente da reunião de cinco outros serviços, a saber: AgriFor<sup>47</sup>: *the Uk's guide to high quality Internet resources in agriculture, food and forestry*, VetGate<sup>48</sup>: *the Uk's gateway to high quality Internet resources in animal health*, OMNI<sup>49</sup>: *the Uk's gateway to high quality Internet resources in health and medicine*, Natural Selection<sup>50</sup>: *the Uk's gateway to high quality Internet resources relating to the natural world* e BIORES<sup>51</sup> (BioResearch): *the Uk's gateway to high quality Internet resources in biological and biomedical research*. Todos estes Gateways contribuem na sua área específica para a construção de um grande catálogo de recursos na área temática das Ciências da Vida.

A apresentação pública do novo serviço deu-se em Novembro de 2002, constituindo o culminar de um processo iniciado em 1999 e protagonizado pela Universidade de *Nottingham* e pelo *Natural History Museum*, aos quais se foram juntando o *CTI Biology* da Universidade de *Liverpool*, o *Royal College of Veterinary Surgeons*, a Universidade de *Oxford*, a Universidade de *Reading*, o *BRIL – BBSRC Research Institute Librarians Organization* e o *Royal Free Hospital*, que constituem actualmente o consórcio responsável pelo serviço, liderado por *Nottingham*.

A construção deste serviço, no seu *layout* geral, nas formas utilizadas para organizar e disponibilizar acesso à informação seleccionada segue de perto a experiência anterior do *Gateway OMNI*, criado em 1995 e da responsabilidade, exactamente, da Universidade de *Nottingham*. O objectivo assumido para este serviço é, aliás, claramente referenciado<sup>52</sup> como sendo a extensão do conceito OMNI (que abarcava a área da saúde e das ciências da vida) a áreas tematicamente associadas.

O processo de selecção, avaliação e descrição dos recursos a incluir no catálogo obedece a políticas comuns claramente definidas que são seguidas por cada serviço. Como teremos oportunidade de verificar, as diferenças registam-se maioritariamente na selecção das linguagens documentais utilizadas para proporcionar o *browsing* dos recursos. Através de uma política de desenvolvimento da colecção<sup>53</sup>, definem-se a abrangência temática do serviço, o público alvo, o tipo de recursos a incluir, as metodologias para assegurar a manutenção dos recursos e, finalmente, as fontes de informação a utilizar para seleccionar novos recursos. A avaliação dos recursos obedece a critérios gerais nas áreas do contexto da produção da informação, do conteúdo e do

---

<sup>46</sup> Disponível em <http://www.biome.ac.uk>

<sup>47</sup> Disponível em <http://agrifor.ac.uk>

<sup>48</sup> Disponível em <http://vetgate.ac.uk>

<sup>49</sup> Disponível em <http://omni.ac.uk>

<sup>50</sup> Disponível em <http://nature.ac.uk>

<sup>51</sup> Disponível em <http://bioresearch.ac.uk>

<sup>52</sup> Cf. BIOME background. Disponível em <http://biome.ac.uk/about/background.html>

<sup>53</sup> Cf. BIOME collections policy. Disponível em: <http://biome.ac.uk/guidelines/collection/>

formato, mas também à definição de aspectos específicos para cada um dos serviços.<sup>54</sup> A descrição dos recursos é realizada de acordo com um conjunto de regras estabelecidas num manual construído para o efeito.<sup>55</sup>

Com a extensão a praticamente todas as áreas da saúde e das ciências da vida, BIOME pretende proporcionar aos seus utilizadores quer a possibilidade de pesquisar através de todos os recursos na área das ciências da vida, quer uma pesquisa focalizada em cada uma das áreas específicas.

A concretização do primeiro objectivo traduz-se na possibilidade de pesquisar por palavra-chave todos os *Gateways* a partir de um mesmo interface. Esta pesquisa pode ser executada em duas opções diferenciadas: a pesquisa simples e a pesquisa avançada. No primeiro caso, o sistema pesquisa por defeito todas as palavras introduzidas pelo utilizador nos campos título, descrição e palavras-chave do registo e assume também por defeito a truncatura de todas as palavras. Ainda nesta primeira forma, o utilizador tem à sua disposição outras opções de refinamento da sua pesquisa, a saber: a pesquisa por frase exacta (introduzida entre aspas), a pesquisa com os operadores booleanos E e OU. Na pesquisa avançada, para além das opções já descritas é ainda possível restringir por tipo de recursos de acordo com uma tipologia de recursos definida conjuntamente para todos os *Gateways* que compõem este serviço, restringir por *Gateway*, ou seja, seleccionar um ou mais *Gateways*, desactivar a truncatura definida por defeito, o que evita para alguns casos um excesso de resultados, e finalmente configurar a visualização dos resultados no que diz respeito a: número total de resultados mostrados, número de resultados mostrados em cada página, critério de ordenação dos resultados (as opções são apenas título ou título dentro de cada *Gateway*), visualização dos metadados (a opção reduz a possibilidade de mostrar apenas os títulos). Quando a pesquisa é executada sobre todos ou vários *Gateways* os resultados são ordenados alfabeticamente pelo título e contém os seguintes elementos de informação: título, descrição e palavras-chave. Cada um dos resultados é acompanhado por uma barra colorida posicionada à esquerda do texto, sendo o sinal distintivo da sua origem em termos de *Gateways*.

A partir da página inicial de BIOME é possível aceder a cada um dos *Gateways* individuais, onde são disponibilizadas opções de pesquisa por palavra-chave que se revestem das mesmas características estruturais da anteriormente descrita.

A opção de *browsing*<sup>56</sup> dos recursos concretiza dificilmente o objectivo de proporcionar uma pesquisa cruzada entre todos os serviços, visto que se reduz apenas à hiperligação que conduz o utilizador a cada uma das estruturas de *browsing* individuais dos serviços. A impossibilidade de proporcionar uma opção que permita o *browsing* cruzado dos vários serviços fica a dever-se à utilização por estes de diferentes linguagens documentais para esse efeito e às dificuldades de implementação de um qualquer processo que operacionalize essa hipótese.

---

<sup>54</sup> Cf. Resource evaluation for BIOME. Disponível em: <http://biome.ac.uk/guidelines/eval/>

<sup>55</sup> Cf. Gray, Lisa – Cataloguing rules for BIOME service: a procedural manual. Disponível em: <http://biome.ac.uk/guidelines/cat/>

<sup>56</sup> Este assunto será objectivo de um tratamento mais desenvolvido, pelo que aqui apenas se faz uma breve descrição.

A partir da página principal de BIOME, o utilizador pode aceder a cada um dos *Gateways* específicos, nos quais encontra o mesmo modelo de organização e recuperação da informação, desta vez sobre uma área temática mais específica, a saber: OMNI – Saúde e Medicina (actualmente disponibiliza 7.103 descrições de recursos<sup>57</sup>)  
Agrifor – Agricultura, Alimentação, Florestas  
VetGate – Saúde Animal  
BioResearch – Ciências biológicas e biomédicas  
Natural Selection: Botânica, Zoologia, Paleontologia, Ecologia, Ciências do Ambiente.

O *Gateway* Agrifor é da responsabilidade da *Greenfield Medical Library* da Universidade de *Nottingham* e destina-se a um público-alvo de estudantes, investigadores, académicos e práticos nas áreas temáticas das quais se ocupa. Cooperam com a biblioteca de *Nottingham* no desenvolvimento deste serviço as universidades de *Oxford* e *Reading*, bem como o *Finnish Forest Research Institute*. Contém actualmente 2.850 recursos descritos.<sup>58</sup>

O *Gateway* Natural Selection é da responsabilidade do *Natural History Museum*, com a cooperação das universidades de *Oxford* e de *Reading*. Destina-se a uma larga faixa de público, desde as crianças em idade escolar até investigadores seniores. Contém actualmente 3.451 recursos descritos.<sup>59</sup>

O *Gateway* VetGate é da responsabilidade da *Greenfield Medical Library* da Universidade de *Nottingham* e destina-se a um público de estudantes, investigadores, académicos e práticos nas áreas temáticas das quais se ocupa. Com a biblioteca cooperam ainda o *Royal College of Veterinary Surgeons* e o *Royal Veterinary College*. Contém actualmente 2.016 recursos descritos.<sup>60</sup>

O *Gateway* BioResearch destina-se igualmente a um público maioritariamente universitário nas áreas temáticas das quais se ocupa e é também da responsabilidade da *Greenfield Medical Library*. Contém actualmente 2.256 recursos descritos.<sup>61</sup>

---

<sup>57</sup> Dados retirados em 03/02/24

<sup>58</sup> Dados retirados em 03/02/27

<sup>59</sup> Dados retirados em 03/02/22

<sup>60</sup> Dados retirados em 03/02/24

<sup>61</sup> Dados retirados em 03/01/27

## Cap. IV – Organização da Informação nos *Subject Gateways* da RDN

### 4.1. Metodologia de análise dos processos e formas de organização da informação

Tendo em conta o quadro conceptual (pontos 2.3. e 2.4. especialmente), as linguagens de metadados dos serviços serão analisadas de acordo com os seguintes critérios:

1. Tipo de Linguagem, considerando dois grandes tipos: Linguagem da **Obra** e Linguagem do **Documento**, enquanto invólucro da informação
2. Componentes da linguagem, nomeadamente entidades a descrever, atributos dessas entidades, vocabulário utilizado para designar entidades e seus atributos, semântica, sintaxe e relações.

A aplicação dos critérios será realizada através de uma análise de conteúdo das regras de catalogação / descrição definidas para cada serviço a partir de uma grelha de categorias que concretizam os critérios definidos, como se segue:

**Tabela 10 – Grelha das categorias de análise das linguagens de metadados dos serviços em estudo**

Critérios		Categorias			
1. Tipos de linguagem		1.1. Atributos da Linguagem da <b>Obra</b>			
		1.2. Atributos da Linguagem do <b>Documento</b>			
2. Componentes	2.1. Entidades	2.1.1. Entidades a descrever			
		2.1.2. Definição das entidades			
2.1.3. Entidades VS Atributos					
	2.2. Atributos da informação	2.2.1. Definição e número de atributos			
		2.2.2. Cardinalidade			
		2.2.3. Granularidade			
	2.3. Vocabulário	2.3.1. Termos autorizados para designar atributos, entidades e relações			
		2.3.2. Regras de forma			
		2.3.3. Tipos de termos: derivados, atribuídos			
		2.4. Semântica	2.4.1. Semântica relacional	2.4.1.1. Relações de sinónimos	
				2.4.1.2. Relações hierárquicas	
				2.4.1.3. Termos relacionados	
	2.4.2. Semântica referencial	2.4.2.1. Formas de limitar o sentido dos termos			
2.4.3. Semântica categorial		2.4.3.1. Facetas ou categorias em que o vocabulário é dividido			
	2.5. Sintaxe	Ordem dos elementos			
Sinais sintácticos					
2.6. Relações	2.6.1. Definição e número de atributos da relação				
	2.6.2. Tipos de relação				

Considerando que qualquer linguagem bibliográfica está ao serviço dos objectivos definidos para o sistema de informação em presença e tendo ainda em conta que o modelo em que estes serviços se baseiam implica a assunção de objectivos “universais” já referenciados, far-se-á num segundo momento uma análise comparativa das

linguagens face a esse referencial. Assim, verificar-se-á a “probabilidade” de consecução dos referidos objectivos através dos seguintes indicadores:

**Tabela 11 – Indicadores da consecução dos objectivos universais dos sistemas de informação bibliográfica**

<b>Objectivos</b>	<b>Indicadores</b>
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título
	Atributo Título alternativo
	Atributo Versão
	Atributo autor (atributos que permitem distinguir autores semelhantes)
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua
	Atributo Formato
	Atributo Requisitos de Acesso
	Atributo Descrição
	Atributo Assuntos
	Atributo Público-alvo
	Atributo Tipologia do recurso
	Atributo Data
	Atributo Versão
	Atributo Cobertura
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: “É versão de” e “Tem versão”
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)
Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)
Obter acesso à entidade descrita	Atributo URL / URN / URI
	Atributo ISBN /ISSN
Navegabilidade no universo bibliográfico	Atributo Relação
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos

A análise das regras da descrição dos vários serviços não contempla, como é evidente, o nível da sua concretização. Assim, a partir de uma amostra de 50 registos de cada serviço seleccionada aleatoriamente será verificado o nível de aplicabilidade e concretização das referidas regras através de uma bateria de critérios que, da mesma forma que no caso da “teoria” permite verificar a consecução dos objectivos universais de um sistema de informação bibliográfico, concretizados através de um conjunto pertinente de indicadores e respectivo campo de aplicação:

**Tabela 12 – Critérios e Indicadores para análise da amostra**

<b>Critérios</b>	<b>Indicadores</b>	<b>Campo de aplicação</b>
Aplicabilidade do nível de exaustividade da descrição	. Nº de elementos por registo . Ocorrência dos elementos obrigatórios . Ocorrência dos elementos	. Todo o registo bibliográfico

	<p>facultativos</p> <ul style="list-style-type: none"> <li>. Ocorrência de elementos em branco</li> </ul>	
<p>Capacidade de discriminação entre recursos semelhantes (encontrar uma entidade singular)</p>	<ul style="list-style-type: none"> <li>. Utilização de complementos de título</li> <li>. Utilização de títulos alternativos</li> <li>. Utilização do atributo Autor e de informação complementar</li> <li>. Ocorrência da prática de desenvolvimento de acrónimos</li> <li>. Ocorrências de erros ortográficos</li> <li>. Ocorrência de erros de formato</li> <li>. Ocorrência de informação relativa a versão.</li> </ul>	<ul style="list-style-type: none"> <li>. Elementos título, autor, data, versão (edição)</li> </ul>
<p>Acessibilidade à entidade descrita</p>	<ul style="list-style-type: none"> <li>. Utilização de URL / URN / URI</li> <li>. Validade das ligações</li> <li>. URL alternativo</li> </ul>	<ul style="list-style-type: none"> <li>. Elemento Endereço do recurso</li> </ul>
<p>Capacidade de identificação e selecção de uma entidade apropriada às necessidades do utilizador</p>	<ul style="list-style-type: none"> <li>. Ocorrência dos atributos: <ul style="list-style-type: none"> <li>- Língua</li> <li>- Formato</li> <li>- Requisitos de acesso</li> <li>- Descrição (legibilidade, validade / qualidade do conteúdo)</li> <li>- Assuntos</li> <li>- Público-alvo</li> <li>- Tipologia do recurso</li> <li>- Data criação / Data última actualização</li> </ul> </li> <li>. Ocorrência da prática de desenvolvimento de acrónimos</li> <li>. Ocorrências de erros ortográficos</li> <li>. Ocorrência de erros de formato</li> <li>. Ocorrência de informação relativa a versão.</li> </ul>	<ul style="list-style-type: none"> <li>. Elementos respectivos ou parte de elementos</li> </ul>
<p>Navegabilidade do universo bibliográfico</p>	<ul style="list-style-type: none"> <li>. Ocorrência do atributo Relação</li> <li>. Ocorrência de ligações entre os recursos, quando se utiliza qualquer uma das instanciações do elemento relação</li> <li>. Ocorrência de ligações nos cabeçalhos de assunto e autor a partir de uma descrição individual</li> <li>. Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos</li> </ul>	<ul style="list-style-type: none"> <li>. Elementos Relação, Assunto, Autor</li> </ul>
<p>Capacidade para localizar conjuntos de entidades: todos os documentos da mesma <b>Obra</b></p>	<ul style="list-style-type: none"> <li>. Ocorrência de títulos uniformes ou ligações transparentes entre formas alternativas do título</li> </ul>	<ul style="list-style-type: none"> <li>. Elemento título</li> </ul>
<p>Capacidade para localizar conjuntos de entidades: todos os recursos da mesma versão</p>	<ul style="list-style-type: none"> <li>. Ocorrência dos qualificadores do elemento relação: "é versão de ", "tem versão"</li> </ul>	<ul style="list-style-type: none"> <li>. Elemento Relação</li> </ul>
<p>Capacidade para localizar conjuntos de entidades: todos os recursos de um mesmo autor</p>	<ul style="list-style-type: none"> <li>. Ocorrência de controlo de autoridade autor ou de outra forma de controlo (ligações transparentes entre formas alternativas dos nomes)</li> </ul>	<ul style="list-style-type: none"> <li>. Elemento autor</li> </ul>
<p>Capacidade para localizar conjuntos de entidades: todos os recursos com o mesmo assunto</p>	<ul style="list-style-type: none"> <li>. Utilização de vocabulário normalizado e controlado.</li> <li>. Ocorrência de controlo de autoridade assunto ou de outra forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)</li> <li>. Ocorrência de estruturas</li> </ul>	<ul style="list-style-type: none"> <li>. Elemento Assunto</li> </ul>

	organizativas que propiciem <i>browsing</i> dos recursos	
--	--	--

Os atributos definidos como indicadores para cada critério são aqueles que mais especificamente contribuem para a verificação dos objectivos, o que não exclui evidentemente o papel mais ou menos relevante que todos eles desempenham para a concretização de todos os objectivos.

Por último e considerando a relevância para efeitos de recuperação da informação de todos os aspectos relacionados com a organização da informação a partir do seu conteúdo, os atributos **ASSUNTO** (Palavras-chave e Classificações), **DESCRIÇÃO** e as **ESTRUTURAS DE BROWSING** foram objecto de uma análise mais profunda, utilizando critérios de maior especificidade, de forma a complementar os que foram definidos para estes domínios na análise global das linguagens da descrição.

Os atributos **classificação** e **palavras-chave** desempenham neste contexto um duplo papel: são atribuídos individualmente a cada descrição, servindo como pontos de acesso a essa descrição, sendo utilizados em seguida para gerar as estruturas de *browsing* para os recursos, dependendo a sua utilização na totalidade (classificações e palavras-chave) da decisão individual de cada serviço. Assim, a análise dos dois aspectos far-se-á nesta dupla vertente: enquanto atributos da descrição e na sua concretização nas estruturas de *browsing* dos recursos.

Considerando a abordagem relativa à representação do conteúdo em metadados (pontos 2.5.2. e 3.2.1), definiram-se os seguintes critérios de análise, aplicados à amostra:

1. Linguagens utilizadas:
  - 1.1. Vocabulários controlados / não controlados
  - 1.2. Tipo de linguagens: classificações, tesouros, listas de cabeçalhos
  - 1.3. Utilizações de diferentes linguagens no mesmo serviço
2. Selecção das linguagens, sua adequação às temáticas do serviço e ao público-alvo
3. Formas de complementaridade entre as linguagens
  - . Classificações – generalidade
  - . Tesouros – especificidade
4. Implementação das linguagens
  - 4.1. Exaustividade e especificidade da indexação / classificação
  - 4.2. Consistência da indexação / classificação
  - 4.3. Tratamento de sinónimos
  - 4.4. Escrita dos termos
  - 4.5. Relação classificações / tesouros
  - 4.6. Ligações entre termos: a partir de um termo aceder a:
    - Termo hierarquicamente superior
    - Termo hierarquicamente inferior
    - Termo relacionado
5. Utilização de outros elementos associados à representação do conteúdo<sup>62</sup>:
  - Cobertura (espaço / tempo)

Na medida em que as classificações são utilizadas para organizar as funcionalidades de *browsing* dos recursos, ou seja, os termos das classificações atribuídos em cada

<sup>62</sup> Excluem-se aqui as descrições em texto livre que serão abordadas especificamente.

descrição são utilizados para posicionar os recursos na estrutura de *browsing*, é neste contexto que estas linguagens serão analisadas.

Tendo em conta o quadro teórico de abordagem do *browsing* dos recursos (Cf. Ponto 2.5.3), definiram-se os seguintes critérios de análise, aplicados a cada uma das estruturas de *browsing* dos serviços analisadas na sua totalidade<sup>63</sup>:

- 1- Organização da estrutura de *browsing*, ou seja, que forma é seleccionada para a organização dos termos utilizados (alfabética, hierárquica, associativa, outra) e que tipo de linguagens documentais para a representação do assunto são utilizadas
- 2- Nível de especificidade na representação dos assuntos. Que classes e sub-classes são utilizadas e a que nível de especificidade são os assuntos representados dentro dessas classes. Esta análise só se torna pertinente no caso das classificações.
- 3- Distribuição dos recursos classificados nos diversos níveis da estrutura. Pretende-se analisar de que forma são atribuídos os recursos nos diversos níveis da estrutura de assuntos, de forma a detectar diferentes abordagens eventuais relativamente à especificidade da atribuição, e verificar a lógica das subdivisões estruturais.
- 4- Contexto. As linguagens documentais proporcionam, de uma forma ou de outra (hierárquica, associativa) a contextualização dos termos utilizados.
- 5- Notações. No contexto das classificações, a utilização das respectivas notações pode trazer vantagens na medida em que estas são independentes da língua do descritor associado e para o caso das classificações tradicionais há que contar com a possibilidade de utilizadores já habituados ao seu significado.
- 6- Atribuição. Uma das principais vantagens da utilização das classificações neste contexto, é a possibilidade de atribuir o mesmo recurso a várias classes ou sub-classes. A análise far-se-á através de uma medida que designamos por taxa de atribuição calculada da seguinte forma: 
$$\frac{\text{Número de recursos}}{\text{Número de recursos no } \textit{browsing}} \times 100$$
- 7- Escrita dos descritores. Devem excluir-se abreviaturas ou truncatura, por dificultarem a compreensão do utilizador.
- 8- Características significativas do interface para o utilizador:
  - 8.1- Transparência da estrutura
  - 8.2- Presença de informação sobre o caminho percorrido na navegação
  - 8.3- Indicação das relações entre os termos
  - 8.4- Informação sobre o número de recursos em cada categorias e/ou sub-categoria

A partir da abordagem conceptual das descrições em texto livre (Cf. Ponto 2.5.4), estas serão analisadas quer do ponto de vista teórico, quer prático: no primeiro caso, através de uma análise de conteúdo do atributo **descrição** em cada uma das regras de catalogação dos serviços; no segundo através de uma análise da amostra que visa avaliar a sua legibilidade. A avaliação da legibilidade será realizada através da submissão da amostra aos seguintes testes de legibilidade utilizados por outros estudos do mesmo tipo: número de palavras, número de caracteres, número de parágrafos, número de frases, média do número de frases por parágrafo, média do número de palavras por

---

<sup>63</sup> Uma metodologia semelhante a esta foi utilizada por Diane Vizine-Goetz no estudo "Classification schemes for Internet resources revisited", in High-Level subject access tools and techniques in Internet cataloging. New York: Haworth Press, 2202, pp. 5-18

frase, média do número de caracteres por palavra, percentagem das frases na voz passiva, Flesh Reading Ease e Flesh Kincaid Grade Level.

Excluíram-se da análise todos os elementos de metadados relacionados com a administração do recurso, considerando que a perspectiva de abordagem deste trabalho é a da descrição com objectivos de recuperação da informação.

## 4.2. Linguagens de metadados

A definição de uma linguagem para a descrição apresenta, como já tivemos oportunidade de verificar, inegáveis vantagens e é condição essencial para garantir a validade destes serviços. A análise das linguagens construídas por cada um dos serviços constitui, portanto, na economia deste estudo, um dos aspectos mais importantes.

As opções de cada um dos serviços na construção das suas linguagens para a descrição têm um significado particular na medida em que nelas transparecem as perspectivas conceptuais que orientam a construção destes universos.

### 4.2.1. SOSIG

As regras de catalogação<sup>64</sup> deste serviço são as que mais se aproximam do normativo definido para ROADS, quer no que diz respeito às entidades a descrever, quer aos atributos definidos para essas entidades, quer ainda relativamente aos elementos de administração. Outra fonte para a elaboração é o já referido manual de Rebecca Bradshaw elaborado para o serviço ADAM.

As entidades objecto de descrição (Cf. Tabela 13) são a **Obra**, o **Documento**, o **Assunto**, mais especificamente a representação do conteúdo através de uma linguagem de tipo classificatório, e o que poderíamos designar por **Responsabilidade**, subdividida em autor enquanto indivíduo ou indivíduos responsáveis intelectualmente pelo conteúdo e organizações (autores pessoas colectivas). As entidades que expressam qualquer tipo de responsabilidade perante o recurso, bem como a entidade assunto partilham uma dupla natureza: a de serem entidades em si próprias mas também atributos da entidade **Obra**. Incompreensivelmente o atributo **palavras-chave** não recebe o tratamento de entidade independente tal como o atributo assunto, tanto mais que as palavras-chave têm uma única origem num vocabulário controlado definido para este serviço (Cf. Ponto 4.3.1.)

Ao preverem a existência de diversos *templates* (no âmbito do pacote aplicativo ROADS), as Regras de Catalogação em análise definem uma série de entidades operacionais de acordo com vários critérios: forma de estruturação da informação diferenciando “Documento” enquanto uma peça individual de informação, de “Serviço” enquanto colecção de informação com hiperligações para outros objectos; tipo de media, distinguindo “Som”, “Vídeo” e “Imagem”; tipos de conteúdos, diferenciando “Conjuntos de Dados” e “Programas” e por último, entidades informativas que resultam

---

<sup>64</sup> Toda a análise deste serviço se baseia em HIOM, Debra – Sosig cataloguing rules, versão 3.0, 1999, pelo que se suprimem todas as referências directas a este documento ao longo do texto.

das especificidades de comunicação na rede diferenciando “Arquivos de listas de correio electrónico” e “Grupos de Discussão”.

No entanto, só as entidades “Documento” e “Serviço” são verdadeiramente consideradas ao longo da definição dos atributos, já que se verifica uma completa ausência de atributos específicos para os outros tipos definidos. Assim, os atributos específicos da entidade Serviço são: Política de acesso, Tempos de Acesso, Autenticação, Política de Preços, Registo, Tamanho, excluindo-se também os seguintes: Tipo de caracteres, Citação, Direitos de Autor, Data de Criação, Formato, Data da última revisão, Estado de publicação, Versão. No caso da entidade **Documento** apenas se especifica que o atributo Autor só pode ser utilizado como atributo desta entidade e não com a entidade Serviço.

As especificidades atribuídas ao Serviço dizem apenas respeito à entidade **Documento** enquanto invólucro físico, já os atributos excluídos contêm um pouco inexplicavelmente os atributos Citação e Versão, sobretudo mais inexplicável no segundo já que se considerarmos as definições de **Documento** e **Serviço** a probabilidade da existência de versões da mesma entidade informativa é, pelo menos, equivalente. O resto dos atributos excluídos apresentam alguma consistência com o tipo de recursos a descrever nesta categoria, já que são sobretudo considerados sítios *Web* institucionais, o que explica a inexistência de direitos de autor, do tipo de caracteres visto só se preverem sítios originários da civilização ocidental, das datas tendo em conta a constante alteração prevista, do estado de publicação não sendo esperável que um sítio organizacional seja publicado “em construção”, do formato já que se trata sempre de objectos hipertextuais, do autor pessoa física pela origem institucional dos sítios. No entanto, esta diferença entre **Documento** e **Serviço** apresenta alguma ilogicidade vista à luz das entidades do universo informativo tal como as temos vindo a identificar, senão vejamos: qual a diferença entre um sítio *Web* e um documento que contenha hiperligações da responsabilidade de um autor individual? qual a diferença entre um determinado autor pessoa física publicar uma bibliografia (categoria incluída no conceito de **Documento**) de recursos *Web* ou produzir um sítio *Web* com uma bibliografia organizada de acordo com uma dada temática?, qual a diferença entre um autor pessoa física publicar material educativo (categoria incluída no conceito de serviço) ou uma organização? As diferenças prendem-se verdadeiramente mais com os produtores do que com outro qualquer critério. Esta tendência é perceptível pela análise das categorias definidas para cada entidade. Vejamos alguns exemplos: “*Articles/Papers/Reports (individual)*” como categoria da entidade **Documento** por oposição a “*Articles/Papers/Reports (collections)*” na entidade Serviço; “*Bibliographies (individual lists of bibliographic information, not contained within a database)*” por oposição a “*Bibliographic databases*”. Esta distinção não está isenta de um certo número de contradições, nomeadamente a consideração de organizações enquanto produtoras de informação para entidades informativas definidas como unidades singulares, como é exemplo a categoria “*Government Publications (online documents published by government bodies)*” inscrita na entidade **Documento** enquanto uma categoria intimamente relacionada “*Government bodies (Web sites produced by government and government bodies)*” instanciação possível da entidade Serviço.

Por outro lado, apesar desta tentativa de discriminação das entidades, a questão da granularidade da descrição não fica completamente resolvida. Questões como, por exemplo, a de saber que nível de especificidade de descrição deve ser empregue no caso

de um sítio *Web* de uma organização governamental que contenha uma publicação governamental, ou de um sítio *Web* que contenha uma área de questões frequentes continuam por resolver.

Considerando desta feita os atributos no seu conjunto (Cf. Tabela 13) verifica-se que, em termos quantitativos, se observa um certo desequilíbrio entre as entidades **Obra** e **Documento**, já que para a primeira são previstos 11 atributos e para a segunda 19. As entidades ligadas à responsabilidade são descritas através do mesmo número de atributos (8), exceptuando a entidade organização com menos um atributo (7). A entidade assunto é descrita com apenas dois atributos. O número de atributos possíveis é, assim, de 56.

Os atributos da entidade **Obra** desenvolvem-se em torno das seguintes características essenciais:

- a) **Títulos**, contemplando três tipos de título: **título próprio**, **título alternativo** e **complemento de título** (embora não previsto formalmente como atributo, a sua existência depende-se das referências feitas aquando da definição do título próprio)
- b) **Assuntos**, prevendo-se atributos para dois tipos de linguagens: classificações e tesouros. Ainda claramente dentro desta área da representação do conteúdo encontramos o atributo Descrição.
- c) **Responsabilidades**, considerando-se dois tipos de autores: **autores pessoas físicas** e **autores pessoas colectivas** (organizações). A noção de diferentes tipos de responsabilidade segundo a importância do papel desempenhado por cada responsável está ausente, não havendo portanto lugar à distinção entre responsabilidades principais e secundárias.

De entre os restantes atributos destaca-se o atributo **fonte** por ser o único que pode denotar a ideia de relação entre recursos. No entanto, a sua definição revela uma importante limitação face à diversidade de tipos de relações possíveis entre obras neste ambiente já que prevê apenas um tipo de relação com o mesmo conteúdo em suporte papel.

Os atributos que descrevem o invólucro do conteúdo apresentam uma diversidade significativa visando cobrir vários aspectos particulares dos “documentos” produzidos neste ambiente, no que diz respeito a acessibilidade, publicação e características físicas. As responsabilidades relativas à publicação no sentido de disponibilizar publicamente o recurso e garantir a sua continuidade ao longo de tempo são particularmente alargadas a funções para além do editor “tradicional”, tentando abarcar os diversos agentes possíveis que podem ter um papel a desempenhar na disponibilização desse recurso, nomeadamente Detentores do Recurso, Administradores do Recurso, Patrocinadores do Recurso, podendo, de facto, estes vários papéis ser desempenhados por organizações diferentes. Os atributos definidos para cada uma destas entidades visam claramente identificá-los de forma a mais inequívoca possível bem como fornecer informação que permita validar a sua qualidade. Estas preocupações prendem-se evidentemente com as novas características de publicação neste ambiente, características essas que determinam também o atributo **estado de publicação**. Os atributos relativos à acessibilidade são particularmente desenvolvidos no que respeita às condições de utilização (Cf. Tabela 13), prevendo-se igualmente URLs alternativos no caso de recursos em outras línguas. As características físicas dos objectos são descritas através de atributos que pretendem

dar a sua dimensão física, tentativa que pode parecer de difícil concretização para um atributo como **tamanho**, mas já mais exequível para atributos como **tipo de caracteres** e **formato**.

Os atributos previstos para descrever os autores são exaustivos (Cf. Tabela 13) e pretendem assegurar a validade da sua qualidade de autor, bem como distinguir o mais inequivocamente possível autores com onomástica semelhante.

Os atributos para descrever o Assunto estipulam, para além do descritor da notação, a identificação da linguagem documental de onde esse assunto é retirado.

A maior parte do vocabulário a utilizar é derivado da fonte prescrita de informação, sendo que este conceito nunca é claramente definido, entendendo-se, pela análise das regras, que será o recurso no seu todo. O vocabulário atribuído controlado é definido para os seguintes atributos: **língua, país, categoria do recurso, conjunto de caracteres, datas**, quer através de normalização interna ao serviço (caso da categoria do recurso), quer pelo recurso a normalização internacional. Só num caso o nível de controlo não é explícito: trata-se do atributo **fonte** onde se estabelece que a referência à fonte deve conter os “detalhes bibliográficos apropriados” sem se indicar qual a regra de referência bibliográfica a utilizar. A utilização de vocabulário atribuído mas sem controlo é prevista em vários casos, em situações em que não é possível transcrever o conteúdo do atributo da fonte prescrita de informação. A utilização de maiúsculas e minúsculas é estipulada em algumas situações, nomeadamente no caso de palavras-chave (minúsculas) e na capitalização dos nomes próprios. Por último, refira-se que no caso dos atributos **assunto** e **palavras-chave** o controlo vocabular é completo, pelo recurso a linguagens absolutamente normalizadas (ver ponto 4.3.1)

O único aspecto normativo relativamente à semântica dos atributos é de tipo referencial para as entidades **Autor, Administrador** e **Organização**, permitindo através de um extenso conjunto de atributos distinguir entidades semelhantes. O mesmo acontece com a entidade **Assunto** através do atributo **esquema do descritor de assunto**

São definidas algumas regras sintáticas, estipulando-se a ordem dos elementos na frase para o caso dos autores, onde se prevê a entrada inversa, e para casos como endereço do local de trabalho de autores e administradores. Os sinais sintáticos são utilizados para distinguir atributos relacionados (distinção entre títulos próprios e complementos de título) ou a repetição de conteúdos de um mesmo atributo (por exemplo, várias palavras-chave separadas por vírgula). No caso dos atributos de responsabilidade, o sinal sintático vírgula é utilizado para separar a palavra de ordem de outro elemento do nome.

A cardinalidade dos atributos também não é explicitamente definida, sendo possível pela análise das regras supor a inexistência de limites. No entanto, fica por esclarecer se a repetição do atributo se faz dentro da mesma etiqueta / campo ou se cada repetição dará lugar a novo campo.

Determinados atributos assumem um estatuto de maior importância ao serem considerados obrigatórios, a saber: **categoria do recurso, título próprio, descrição, palavras-chave, administrador – endereço electrónico, assunto – descritor, URI e língua**. A maioria destes elementos diz respeito ao conteúdo intelectual (6 de 8) e

apenas dois ao **Documento** enquanto entidade física. Assim, uma descrição mínima identifica atributos que se compõem maioritariamente de vocabulário atribuído e controlado. A exclusão do atributo **autor** parece não fazer grande sentido tendo em conta o tipo de recursos a incluir no serviço. Tal só eventualmente se justificará pelas dificuldades em obter este tipo de informação no contexto da informação em rede.

Do ponto de vista da capacidade desta linguagem definida para SOSIG de concretizar os objectivos universais de qualquer sistema de informação bibliográfica, verifica-se (Cf. Tabela 14) uma insuficiência significativa no que diz respeito ao objectivo Navegação no universo bibliográfico já que não são consideradas relações entre entidades desse universo, para além daquelas que são possíveis de exprimir pelo atributo assunto. O atributo Fonte tal como é definido revela-se extremamente parco para dar conta de todo o tipo de relações possíveis entre as entidades de informação no universo da rede.

Objectivos como Localizar todos os recursos da mesma **Obra** e Localizar todos os documentos da mesma versão não são claramente considerados, quer pela inexistência dos atributos, quer pela restrição na definição do atributo versão. No caso do objectivo Localizar todos os recursos de um mesmo autor, nada nas regras indica a existência de controlo de autoridade, pelo que podemos concluir igualmente pela sua não verificação.

O objectivo Obter acesso à Entidade descrita é aquele que é mais completamente conseguido.

No que diz respeito ao objectivo localizar todos os recursos sobre um mesmo assunto, não é possível descortinar nas regras qualquer forma de controlo de autoridade, pelo que podemos concluir pela sua deficiente concretização, já que o indicador vocabulário controlado está presente.

Os objectivos com maior probabilidade de serem atingidos são, para além do já referido Obter acesso á entidade descrita, os de Encontrar uma entidade singular e Identificar e Seleccionar uma entidade apropriada às necessidades do utilizador. No primeiro caso apenas as restrições ao atributo **versão** justificam alguma incerteza. No segundo caso, apenas não se verificam completamente os atributos **cobertura** e **público-alvo**.

A análise da amostra revela descrições com uma média de 7 atributos presentes, apresentando uma frequência relativa de 100% todos os atributos obrigatórios, à excepção do atributo endereço electrónico da entidade administrador. De entre os atributos facultativos utilizados, o **título alternativo** é o que apresenta uma maior frequência relativa (34%), seguida do **autor** (10%), **complemento de título** (8%) e URL alternativo (6%).

O atributo **título alternativo** é utilizado para as seguintes situações: títulos em outras línguas que não o inglês, desenvolvimento de acrónimos e outros títulos. Nos dois primeiros casos, a consistência de utilização não é absoluta. Verifica-se por exemplo a utilização do título alternativo *Comunidad Andina* para o título próprio *Andean Community*, bem como o título alternativo *French Association for Balkan Studies* para o título próprio *Association Française d'Études sur les Balkans*, para a primeira situação. Verifica-se igualmente o título alternativo *Search Company Reports* para o título próprio *SCORE*, mas igualmente o título alternativo *DfES* para o título próprio *Department of Education and skills*, na segunda situação. Verifica-se, finalmente, um

caso em que o acrónimo é desenvolvido no título próprio: *IBSS (International Bibliography of the Social Sciences) on line*.

O atributo **autor** é utilizado de forma muito inconsistente: ora se verifica ora é inexistente mesmo em situações onde o autor é referido no conteúdo do atributo descrição (Cf. Ponto 4.3.3) ou de outros. Vejamos dois exemplos:

- a) Michael Nelson é referenciado como compilador do recurso *Leeds Thai Politics Bibliography*, sem o atributo autor estar presente. Neste caso adiciona-se também à referência de autor a sua identificação profissional.
- b) Govinda Shresthe é referida como autora no atributo título do recurso *A review of case studies related to distance education in developing countries by Govinda Shresth, March 1997*, sem ser utilizado o atributo autor.

Dos atributos para identificação e selecção não se verificam os seguintes: **requisitos de acesso, formato, público-alvo, datas, cobertura e editor**. Todos os outros ocorrem com uma frequência relativa de 100%.

Em termos de navegabilidade verifica-se a inexistência do atributo **relação**, embora ocorram ligações nos cabeçalhos de assunto a partir de uma descrição individual no caso da classificação, mas não das palavras-chave e uma desenvolvida estrutura de *browsing* (Cf. Ponto 2.4.3).

A capacidade de identificar todos os recursos do mesmo autor só se verifica nos casos em que este é assim identificado no atributo respectivo e a pesquisa pode realizar quer por entrada directa quer inversa que o resultado é coincidente.

A identificação de todos os documentos do mesmo assunto é apenas conseguida pela utilização de vocabulário controlado (CF. Ponto 4.3.1) e pela existência de uma estrutura de *browsing* (Cf. Ponto 4.3.2). O facto de não ser possível pesquisar num campo de assunto (classificação ou palavras-chave) inviabiliza um teste esclarecedor à hipótese da existência de controlo de autoridade.

À acessibilidade à entidade descrita é assegurada pela maioritária manutenção de ligações válidas entre a descrição e o recurso.

A probabilidade de um utilizador encontrar uma entidade singular encontra-se diminuída pela reduzida importância e tratamento pouco consistente do atributo autor. A identificação e selecção podem realizar-se com recurso a um número relativamente reduzido de atributos. A possibilidade de navegar no universo bibliográfico considera apenas a relação através do atributo assunto. A probabilidade de encontrar conjuntos é muito baixa relativamente a autores e mais provável para os assuntos, embora no que diz respeito à pesquisa por palavra-chave se verifiquem grandes discrepâncias relativamente à pesquisa no tesouros. Localizar todos os recursos da mesma **Obra** e da mesma versão são possibilidades inexistentes.

#### **4.2.2. PSIGATE**

Este serviço baseou a construção da sua linguagem nas regras de catalogação (PSIGATE cataloging rules:2002) para os serviços ADAM e BIOME, nos

procedimentos de administração da base de dados definidos para EEVL e, finalmente, nas regras de catalogação definidas para a RDN.

As entidades para as quais são identificados atributos são **Obra**, **Documento**, **Autor**, **Editor**, **Administrador do recurso** e **Assunto**. Verifica-se uma clara predominância dos atributos relativos à **Obra**, num total de 9, enquanto que para o **Documento** são apenas previstos três atributos. (Cf.Tabela 15)

Os atributos relativos à **Obra** desenvolvem-se maioritariamente à volta de três grandes aspectos: assunto, título e responsabilidade, ou seja, atributos que caracterizam intelectualmente o recurso e o seu conteúdo.

Os atributos relativos ao **Documento** enquanto invólucro da **Obra** são relativamente escassos e prendem-se somente com a localização, um atributo essencial, e com a publicação, definindo entidades para dois tipos de responsabilidade sobre a disponibilização e manutenção do recurso. Essas entidades são apenas descritas pelos atributos nome e endereço electrónico, e constituem igualmente atributos da entidade **Documento** e entidades descritivas em si próprias.

A entidade **Autor** é apenas descrita pelos atributos nome e endereço de correio electrónico, o que é manifestamente insuficiente quer para distinguir autores semelhantes, bem como para determinar a qualidade do seu estatuto.

A entidade **Assunto**, apenas prevista para o caso das classificações, é descrita através de dois atributos, permitindo não só especificar a notação e descritor utilizado mas também o instrumento específico utilizado para essa atribuição. Estranhamente o atributo **palavras-chave** não é considerado como entidade, o que seria tanto mais compreensível quanto a escolha das palavras-chave tem também origem num vocabulário controlado específico.

Em termos do vocabulário a utilizar, ele é sobretudo derivado da fonte prescrita de informação, aqui considerada de forma muito abrangente. Só para três atributos o vocabulário é claramente atribuído e obedece a controlo: as **datas**, a **língua** e a **categoria**. Como é evidente, os atributos **palavras-chave** e **assunto** são igualmente compostos por vocabulário atribuído e controlado, no primeiro caso a partir de um dos Tesouros em uso no serviço e no segundo a partir da CDD21 (Classificação Decimal de Dewey – século 21). No atributo título é igualmente possível utilizar um vocabulário atribuído, na circunstância em que não seja possível identificar claramente um título, mas sem qualquer tipo de controlo. Ainda neste contexto, são estatuídas algumas regras de forma, sobretudo relativamente à utilização das maiúsculas (regras para a capitalização dos títulos e nomes próprios) e à escrita, apontando-se neste caso para a observância do original exceptuando o caso do Editor, para o qual se obriga à utilização da forma de citação deste definida nas AACR2.

Em termos semânticos, são praticamente ausentes quaisquer requisitos ou normas, se exceptuarmos a existência do atributo endereço electrónico para todas as responsabilidades, o que permitirá distinguir nomes semelhantes, se bem que seja um elemento relativamente parco para esse efeito.

Não existem praticamente regras relativas à sintaxe dos elementos, se exceptuarmos a prescrição de separar as palavras-chave por vírgulas e a ordem dos elementos para as datas que estipula a seguinte forma DD/M/AAA. No caso dos nomes, a forma prescrita é a entrada directa. A utilização do espaço tipográfico para a separação de algumas partes dos conteúdos não pode ser exactamente equiparada a uma regra sintáctica.

Não é previsto qualquer atributo capaz de dar a relação entre os recursos.

A obrigatoriedade de alguns atributos (Cf. Tabela 15) não é claramente definida, mas pode ser assumida na medida em que se assinalam os atributos que constituem os requisitos mínimos para a descrição de acordo com as regras de catalogação RDN. Não existem, por outro lado, quaisquer regras relativamente à cardinalidade dos elementos descritivos; ela pode depreender-se para alguns casos quando são estatuídas regras sintácticas ou sinais tipográficos que distinguem conteúdos que se repetem dentro do mesmo elemento.

Do ponto de vista da granularidade da descrição não existe qualquer definição do nível a utilizar. Por um lado, o conceito de recurso Internet nunca é definido. Por outro, a utilização de vários critérios para definir o atributo categoria do recurso, que vão desde a origem institucional do recurso – Departamentos académicos, por exemplo – até tipologias documentais – bibliografias, por exemplo –, passando por tipos de media; bem como a possibilidade de atribuir mais do que uma categoria ao mesmo recurso, não permitem divisar com segurança se a tendência é para descrever o objecto como um todo tal como se apresenta na sua configuração técnica, se partes dele conforme a pertinência tipológica ou de conteúdo.

Esta linguagem revela alguma incapacidade para concretizar os objectivos universais de um sistema de informação bibliográfica, já que só o objectivo Obter acesso à entidade descrita é completamente conseguido. Todos os outros apresentam níveis de consecução diferentes, mas, sobretudo em alguns casos, abaixo do que seria de esperar tendo em conta o modelo implementado.

Apresentando uma total impossibilidade de concretização estão os objectivos Localizar todos os documentos da mesma **Obra**, Localizar todos os documentos da mesma versão e Localizar todos os documentos do mesmo autor, tendo em conta neste último caso a inexistência de qualquer forma de controlo de autoridade prevista (Cf. Tabela 16)

O objectivo Localizar todos os recursos com o mesmo assunto apresenta algum nível de consecução pela utilização de vocabulário controlado, mas a ausência de controlo de autoridade levanta problemas em termos de recuperação podendo inviabilizar assim a possibilidade de reunir todos os recursos com o mesmo tema.

O objectivo Navegabilidade no universo bibliográfico só em parte é atingido. A ausência do atributo relação entre entidades de informação exclui a possibilidade de uma concretização mais efectiva, visto que para além da navegação por via do ponto de acesso Assunto nenhum outro tipo de relação entre os recursos é previsto.

A ausência de vários atributos relevantes para a identificação e selecção de um recurso levanta problemas na concretização deste objectivo nomeadamente no que concerne às

características físicas e de acessibilidade ao recurso, impedindo desta forma o utilizador de saber se reúne as condições tecnológicas e económicas para concretizar o acesso.

O objectivo de encontrar uma entidade singular é o que se apresenta mais próximo de uma completa consecução, na medida em que só relativamente ao atributo **autor** a linguagem apresenta alguma incapacidade de distinguir entidades semelhantes.

A prática descritiva produz descrições com uma extensão média de 7 atributos por descrição, o que corresponde exactamente ao número máximo de atributos mínimos previsto. No entanto, numa análise mais específica verifica-se que nem todos os atributos obrigatórios existindo, ocorrendo, por outro lado, atributos facultativos.

Ocorrem todos os atributos obrigatórios, exceptuando a **língua**. Atributos como **título**, **categoria do recurso**, **descrição**, **palavras-chave** e **URL** apresentam uma frequência relativa de 100%. O atributo **complemento de título** apresenta apenas uma frequência relativa de 20%. Dos elementos considerados não obrigatórios ocorrem a **classificação** (descriptor de assunto) com uma frequência de 100% e **direitos de autor** com uma frequência de 50%.

Verifica-se a existência de uma importante frequência relativa de campos em branco no atributo **direitos de autor**, na ordem dos 50%.

O desenvolvimento de acrónimos ocorre, embora sem frequência absoluta (detectamos pelo menos um caso em que um acrónimo não é desenvolvido).

O sinal sintáctico de distinção entre **título próprio** e **complemento de título** não é utilizado de forma consistente: ora se utilizam os dois pontos ora o ponto. A ocorrência de **títulos alternativos** verifica-se apenas uma vez e num caso de um **título próprio** numa outra língua que não o inglês.

O atributo **autor** nunca é utilizado, apesar da referência a autorias no conteúdo do atributo descrição, como é o caso do Professor William Euler autor do recurso *Advanced organic chemistry I*.

Dos atributos que asseguram a identificação e selecção apenas ocorrem a **descrição**, **assuntos** (palavras-chave e classificação) e **categoria do recurso**, todos com uma frequência relativa de 100%.

Em termos de probabilidade de navegação dentro do universo das descrições verifica-se a ausência do elemento **relação**, bem como da ligação dos cabeçalhos de assunto a qualquer tipo de índice a partir de uma descrição individual. Verifica-se, por outro lado, a existência de uma estrutura para *browsing* dos recursos (Cf. Ponto 4.3.2).

Em termos de acessibilidade, verifica-se a ocorrência de URL a 100%, sendo que apenas quatro ligações se encontram perdidas.

O controlo de autoridade autor é inexistente e o controlo de autoridade assunto, com grande probabilidade, também não se verifica. Dada a inexistência de pesquisa em campos específicos é impossível verificar esta possibilidade completamente. A existência de uma estrutura de *browsing* por assunto pode ser o resultado de um

dispositivo de design do sítio, tanto mais que uma pesquisa ao termo Astronomy (existente na classificação e no *browsing*) apresenta uma diferença de resultados importante entre esta forma de acesso (839 resultados) e o número de recursos identificados no *browsing* para o mesmo termo (1870).

A capacidade de o utilizador encontrar uma entidade singular é apenas possibilitada pela existência de complementos de título, configurando-se portanto uma muito deficiente concretização deste objectivo. A capacidade para localizar todos os documentos da mesma **Obra**, todos os recursos da mesma versão e todos os recursos de um mesmo autor é claramente inexistente. A localização de todos os recursos com o mesmo assunto só é possibilitada pelo emprego de vocabulário controlado (Cf. Ponto 4.3.1) e pela existência de funcionalidades de *browsing* (Cf. Ponto 4.3.2). O acesso à entidade descrita encontra-se integralmente assegurado.

### 4.2.3. BIOME

Todos os serviços que integram BIOME usam a mesma linguagem da descrição (Cataloguing rules for the BIOME service: 2000), pelo que a análise será realizada em conjunto. Esta linguagem tem por base as Regras de Catalogação definidas para o serviço ADAM, as AACR 2 e o manual de Nancy Olson “Cataloguing Internet resources: a manual and a practical guide”.

Definem-se nesta linguagem um extenso conjunto de atributos com função de descoberta e caracterização dos recursos (39) considerando-se as seguintes entidades descritivas: **Obra, Documento, Autor, Organização**. (Cf. Tabela 17). Verifica-se uma clara predominância dos atributos que descrevem os aspectos de produção intelectual, que representam 33.3 % de todos os atributos definidos. Neste domínio, os atributos identificam três grandes características dos objectos de informação: títulos, assuntos e responsabilidade.

Os atributos relativos à responsabilidade são igualmente considerados entidades descritivas, identificando-se, neste caso, um conjunto de atributos que configuram uma caracterização exaustiva da entidade, permitindo não só clarificar a qualidade de autor como distinguir autores com semelhanças onomásticas, revelando o recurso a uma semântica de tipo referencial. (Cf. Tabela 17)

O **Documento** enquanto entidade física é descrito por um conjunto de atributos (cf. Tabela 17) que podem ser agrupados nas seguintes categorias: identificadores, prevendo-se identificadores alternativos quando o “documento” tiver outro endereço ou se encontrar numa outra língua (neste último caso devem utilizar-se os atributos **título alternativo** e **língua**), bem como os identificadores inequívocos tradicionais do ambiente papel; **formato, datas, e estatuto da publicação**. Este último atributo permite ter em conta a evolução editorial do recurso, característica particularmente identificadora das formas de publicação em ambiente *Web*.

Em termos de tipologia vocabular, a grande maioria dos atributos são designados por vocabulário derivado completamente da fonte prescrita de informação, portanto, sem qualquer controlo. O vocabulário atribuído diz apenas respeito aos atributos **datas, categoria do recurso, língua e país**, em todos os casos completamente controlado, quer

através da definição de formatos de ordenação, quer através da definição de códigos que constituem a única possibilidade de designar o atributo (caso da **língua**, **categoria do recurso** e **país**). O único caso de vocabulário atribuído sem controlo diz respeito à possibilidade prevista de atribuição de um título pelo catalogador, quando não esteja presente na fonte prescrita de informação. As regras de forma dizem essencialmente respeito aos nomes que contêm títulos associados, à capitalização de nomes em vários elementos, à predominância de utilização do inglês como língua da descrição e, finalmente, à omissão de artigos definidos e indefinidos no início dos títulos.

A definição de uma sintaxe está praticamente ausente, exceptuando a regra de entrada inversa dos nomes, e o formato para as datas, já que o mero sinal tipográfico para separação de subtítulos não pode ser considerado exactamente como uma regra sintáctica.

A existência do atributo **fonte** é o único caso em que a relação entre recursos é considerada e, mesmo neste caso, apenas se prevê a relação com a versão definitiva desse recurso, hipótese provavelmente pensada para situações em que se trata de transferência de suporte.

Verifica-se a ausência de qualquer regra relativa à cardinalidade dos atributos.

A granularidade da descrição não é definida, visto que nunca se esclarece o conceito de informação objecto do serviço. Por outro lado, também não é possível a partir do atributo categoria divisar qualquer indicação dado que as instanciações desse atributo apontam para uma tipologia documental clássica, sendo relativamente escassas as referências a tipologia própria deste ambiente (tal só acontece nos casos de *Mailing List* e *News Service*)

Do ponto de vista da capacidade desta linguagem definida para BIOME de concretizar os objectivos universais de qualquer sistema de informação bibliográfica, verifica-se (Cf. Tabela 18) uma insuficiência significativa no que diz respeito ao objectivo Navegação no universo bibliográfico já que não são consideradas relações entre entidades desse universo, para além daquelas que são possíveis de exprimir pelo atributo assunto. O atributo Fonte tal como é definido revela-se extremamente parco para dar conta de todo o tipo de relações possíveis entre as entidades de informação no universo da rede.

Objectivos como Localizar todos os recursos da mesma **Obra** e Localizar todos os documentos da mesma versão não são claramente considerados, pela inexistência dos atributos. No caso do objectivo Localizar todos os recursos de um mesmo autor, nada nas regras indica a existência de controlo de autoridade, pelo que podemos concluir igualmente pela sua não verificação.

O objectivo Obter acesso à Entidade descrita é aquele que é mais completamente conseguido.

No que diz respeito ao objectivo Localizar todos os recursos sobre um mesmo assunto, não é possível descortinar nas regras qualquer forma de controlo de autoridade, pelo que podemos concluir pela sua deficiente concretização.

Os objectivos com maior probabilidade de serem atingidos são, para além do já referido Obter acesso á entidade descrita, os de Encontrar uma entidade singular e Identificar e Seleccionar uma entidade apropriada às necessidades do utilizador. No primeiro caso apenas as restrições ao atributo Versão justificam alguma incerteza. No segundo caso, apenas não se verificam completamente os atributos Cobertura e Público-alvo.

De entre todas as descrições produzidas pelos serviços em análise, estas são as que apresentam um menor número de atributos a ser visualizados pelos utilizadores, apenas 3 em média: título próprio, descrição e assunto (palavras-chave). Estes atributos ocorrem sempre, não existindo qualquer campo em branco.

Em relação ao atributo **título**, os **complementos de título** ocorrem em 14% das descrições e os títulos paralelos apenas uma vez.

Os endereços dos recursos não são visualizados, encontram-se “escondidos” na hiperligação a partir do título. Todas as ligações estão activas, exceptuando um caso no serviço Biores.

Os testes realizados para o controlo de autoridade assunto não foram conclusivos, dado que, como não está disponível a pesquisa por campos específicos, torna-se difícil avaliar com eficácia. No entanto, foi possível verificar que a classificação não pode ser pesquisada por palavra, e em relação às palavras-chave a recuperação apenas apresenta um resultado positivo quando se opta pela pesquisa por frase exacta, o que parece indicar a inexistência de controlo de autoridade. No entanto, está disponível a funcionalidade de navegação nas palavras-chave a partir de uma descrição individual, podendo esta implementação ser apenas o resultado de um dispositivo de design hipertextual.

Os resultados visíveis para o utilizador apresentam assim várias dificuldades na concretização dos objectivos básicos. O único objectivo com possibilidade de ser completamente atingido é o de acessibilidade à entidade descrita. No que diz respeito à capacidade para encontrar uma entidade singular, verifica-se apenas a ocorrência significativa do atributo complemento de título. A identificação e selecção de uma entidade apropriada às necessidades do utilizador pode apenas socorrer-se dos atributos descrição (Cf. Ponto 4.3.3) e assunto. A possibilidade de encontrar todos os recursos sobre um mesmo assunto ocorre graças à utilização de vocabulário controlado (Cf. Ponto 4.3.1) e à existência de funcionalidades de *browsing* dos recursos (Cf. Ponto 4.3.2)

Em relação ao estipulado nas Regras de Catalogação deste serviço, a informação descritiva visualizada representa uma drástica diminuição das possibilidades de o utilizador operacionalizar os objectivos mínimos.

#### **4.2.4. EEVL**

De todos os serviços em análise, este é o único que não disponibiliza as regras da descrição utilizadas. Assim, a análise terá como base as referências retiradas de fontes secundárias (DEMPSEY:1997) onde este aspecto é referido, mas sem o

desenvolvimento necessário para aplicar completamente o esquema de análise que temos vindo a utilizar.

A linguagem relativa à **Obra** (Cf.Tabela 19) enquanto conteúdo intelectual organiza-se em volta dos seguintes atributos principais: títulos, representação do conteúdo, tipologia dos recursos e responsabilidade, verificando-se a completa ausência de atributos relativos à relação entre entidades. A linguagem relativa ao **Documento** estabelece atributos relacionados com acessibilidade, especificando condições de utilização e localização, sem quaisquer referências às características físicas ou à publicação.

Verificando-se a inacessibilidade à linguagem deste serviço, a análise da consecução dos objectivos dos sistemas de informação bibliográfica, que se espelha na Tabela 20, tem um valor muito relativo.

Verifica-se, de qualquer forma, que uma insuficiência de concretização de vários objectivos, sendo aqueles que apresentam uma maior probabilidade de sucesso Identificar e Seleccionar, Localizar todos os recursos com o mesmo assunto e Obter acesso à entidade descrita.

A análise da amostra deste serviço revela uma descrição relativamente extensa, com uma média de 10 (9.5) atributos descritos em cada.

Os atributos que verificam a 100% são: **título próprio, URL, descrição, classificação, tipo de recurso, localização, país e língua.**

Os **títulos alternativos** ocorrem em 58% das descrições e são utilizados para situações de desenvolvimento de acrónimos, títulos em outra língua que não o inglês, o mesmo título com outra grafia, e título diferente do título próprio. A alta percentagem de ocorrência deste atributo revela, desta forma, a preocupação em contribuir para distinguir entidades semelhantes e proporcionar uma identificação da entidade mais precisa. Os **complementos de título** ocorrem numa percentagem mais reduzida (14%) e o sinal sintáctico de separação não é usado de forma consistente, já que ora se emprega o hífen, ora os dois pontos.

O atributo **autor** é apenas utilizado em 16% das descrições, consistentemente representado em termos sintácticos através da entrada inversa, mas sem qualquer elemento complementar de identificação.

Dos atributos definidos como indicadores para assegurar um identificação e selecção eficaz só estão presentes a 100% os seguintes: **língua, descrição, classificação, tipologia do recurso.** Os atributos **palavras-chave** e **editor** apresentam percentagens de ocorrência mais reduzidas, respectivamente 54 e 10%. A frequência relativa do atributo **palavras-chave** menor que 100% é um resultado não esperado já que este tipo de atributo é normalmente utilizado de forma completamente consistente.

Em termos de acessibilidade, verifica-se a utilização consistente do **URL**, sendo que apenas num caso se verifica uma alteração de endereço não identificada na descrição, mas proporcionando o respectivo redireccionamento. São utilizados também os identificadores unívocos **ISBN** e **ISSN**, embora ocorram em frequências relativas muito

reduzidas, respectivamente 4 e 2%. URL's alternativos ocorrem numa percentagem de 8%.

No que concerne à possibilidade de navegação no universo bibliográfico não se verifica a existência do atributo **relação**, nem qualquer ligação nos dos cabeçalhos autor e assunto nas descrições individuais para qualquer índice de autor ou assunto. Ocorrem no entanto estruturas que permitem o *browsing* dos recursos (Cf. Ponto 4.3.2).

Os testes realizados para a verificação da ocorrência do controlo de autoridade autor utilizando a pesquisa no respectivo campo, revelaram-se positivos no sentido de que garantem a localização de todos os documentos de um mesmo autor, verificando-se também a independência relativa à forma de entrada.

A possibilidade de localizar todos os documentos sobre um mesmo assunto é assegurada, em parte, pela utilização de vocabulário controlado (cf. Ponto 4.3.1), mas, visto que não é possível pesquisar em nenhum campo específico de assunto (classificação e palavras-chave) é impossível verificar a sua ocorrência.

Não se verificam campos em branco, nem erros ortográficos ou de formato.

Pode concluir-se, então, que a possibilidade de encontrar uma entidade singular se encontra minimamente assegurada, exceptuando pela reduzida ocorrência do atributo autor e sem informação complementar e pela inexistência do atributo versão. A acessibilidade à entidade descrita encontra-se completamente assegurada. A possibilidade de identificar e seleccionar uma entidade apresenta uma probabilidade mais reduzida, visto que se verifica a inexistência dos atributos requisitos de acesso, formato, público-alvo, datas, cobertura e versão. As possibilidades de navegabilidade são apenas proporcionadas pela estrutura de *browsing*. A capacidade de localizar conjuntos de entidades é inexistente para todos os recursos da mesma **Obra** e todos os recursos da mesma versão, conseguida para o caso de todos os recursos do mesmo autor e deficientemente conseguida para todos os recursos com o mesmo assunto.

#### 4.2.5. HUMBUL

A linguagem descritiva preconizada para o serviço HUMBUL, definida num manual próprio actualizado em 2001, apresenta importantes alterações em relação aos casos até agora estudados, desde logo porque a fonte que se baseia é diferente, neste caso Dublin Core. (FRASER:2001)

A linguagem considera atributos para a **Obra**, o **Documento** e o Autor, sendo que a última entidade goza de uma dupla característica: é atributo da entidade **Obra** e entidade em si mesma.

No que diz respeito à linguagem da **Obra** (Cf. Tabela 21) prevêem-se elementos, num total de 18 possibilidades, para atributos nas seguintes áreas: títulos, responsabilidade intelectual, representação do conteúdo (incluindo a cobertura geográfico-temporal), tipologia do recurso, audiência, identidade linguística e relações entre os recursos.

A linguagem do **Documento** define atributos quantitativamente reduzidos, apenas 4, relacionados com acessibilidade, mais especificamente a localização, mas sem prever atributos relativos às condições de acesso, e com a publicação, nomeadamente editor e datas de produção e alteração do recurso.

As linguagens para a responsabilidade intelectual e para o editor apresentam os mesmos atributos: identificação onomástica, função desempenhada, identidade profissional e presença na *Web*. A identidade profissional, dada pelo atributo *Afiliação*, permite ao utilizador reconhecer a validade do produtor enquanto autor, informação particularmente importante no contexto da rede.

Nem todos os elementos têm a mesma importância no contexto da linguagem. Os elementos obrigatórios visam descrever aqueles atributos considerados fundamentais para a caracterização dos recursos, a saber: título, responsabilidade intelectual, localização, publicação, conteúdo, identidade linguística, tipologia Documental.

Grande parte do vocabulário prescrito para o conteúdo dos elementos é vocabulário derivado da fonte prescrita da informação, que pode ser todo o recurso. No entanto, verifica-se uma preocupação em implementar o maior nível possível de controlo vocabular nos elementos em que o vocabulário é atribuído. O controlo vocabular é assegurado por duas vias: pela existência de um conjunto de valores obrigatórios pré-definidos para o conteúdo do elemento – tal é o caso do atributo **função** para autores e editores, do elemento **tipo de recurso**, do elemento **audiência** e do elemento “língua” – ou pela prescrição de um formato obrigatório para a atribuição de um determinado conteúdo – tal é o caso dos atributos **cobertura – período de tempo, cobertura – tempo: data de início / data de fim** e **cobertura – geografia**. Para além destas formas explícitas de controlo, existem outros tipos de regras que sem estabelecer exactamente um vocabulário controlado, indicam soluções para utilização de vocabulário derivado que contribuem para assegurar uma maior consistência de atribuição, por exemplo quando o nome do **autor** não incluir apelido estipula-se o recurso às fontes de referência ou caso o nome de um **autor** seja específico de uma localização geográfica esta deve ser dada entre parênteses, ou ainda a indicação de que se deve considerar edição de autor quando o recurso é disponibilizado por um *Internet Service Provider* generalista sem outra indicação.

Para além do controlo vocabular, outros aspectos de definição semântica estão igualmente presentes: os atributos para autores e editores permitem distinguir entre autores com nomes semelhantes; a prescrição de uma determinada hierarquia para atribuição de responsabilidade no caso de instituições subordinadas permite distinguir entidades subordinadas com designações semelhantes mas que pertencem a uma entidade global diferente.

Em termos de sintaxe, a ordem dos elementos na frase é apenas definida para o caso dos autores, definindo-se a forma invertida da entrada para **autores pessoas físicas** e a forma directa para **autores pessoas colectivas**. São previstos sinais sintácticos para distinguir os elementos nos seguintes casos: dois pontos para separar **títulos** e **complementos de título**, ponto para separação dos níveis hierárquicos dos **autores pessoas colectivas** e, finalmente, vírgula para separar mais do que uma referência dentro do mesmo tipo de **relação**.

As relações entre os recursos são consideradas através de atributos que estabelecem a ligação entre o todo e a parte, a origem do recurso e a ligação referencial entre recursos. Os atributos **recurso requer** e **ligação à declaração de direitos de autor** previstos também no âmbito deste componente da linguagem fogem à noção de relação entre entidades informativas na medida em que o primeiro apenas permite identificar requisitos de acesso (tecnológicos ou outros) e o segundo tem um carácter exclusivamente informativo sobre o recurso, sendo que o facto de se prever uma ligação à declaração de direitos de autor tal não configura qualquer ligação entre recursos. Por outro lado, o significado deste atributo exclui qualquer consideração da relação entre descrições, sempre consideradas como descrições individuais de recursos sem qualquer relação mesmo que se descreva o todo e a parte.

A cardinalidade dos atributos é prevista para o caso dos autores com um máximo de cinco ocorrências e para os editores com um máximo de três.

A granularidade da descrição não é definida.

A capacidade de, através desta linguagem, serem atingidos os objectivos dos sistemas de informação bibliográfica é conseguida praticamente na totalidade para os objectivos Obter acesso à entidade descrita, Navegabilidade no universo bibliográfico (com as limitações já referidas) e Encontrar uma entidade singular (Cf. Tabela 22). No caso do objectivo Navegabilidade se bem que esteja presente o atributo relação, os tipos de relações previstas não contemplam todas as hipóteses (Cf. Tabela 21). A inexistência do atributo versão é o único impedimento à completa concretização do objectivo Encontrar uma entidade singular.

Os objectivos Localizar todos os documentos da mesma **Obra** e todos os documentos da mesma versão são manifestamente impossíveis de atingir.

Os objectivos Localizar todos os recursos de um mesmo autor e todos os recursos de um mesmo assunto falham por via da inexistência de qualquer forma e metodologia de controlo de autoridade explícito, apesar de existir uma estrutura de *browsing* por assuntos que unifica todas as descrições de um mesmo tema.

Os objectivos Identificar e Seleccionar uma entidade apropriada às necessidades do utilizador pode ser atingidos através de um número significativo de atributos previstos. A ausência do atributo **requisitos de acesso** não é completamente verdadeira, na medida em que os conteúdos possíveis deste atributo verdadeiramente existem só que designados pelo atributo **recurso requer**. Assim sendo, os únicos atributos ausentes são **formato e versão**

Analisando, por último, a amostra das descrições seleccionadas de acordo com os critérios definidos verifica-se que o número médio de atributos em cada descrição é de 11.4, oscilando nos valores extremos entre um máximo de 17 e um mínimo de 9, o que vai significativamente para além do que neste contexto poderia ser definido como um registo mínimo, ou seja, o que cumprisse os requisitos dos atributos obrigatórios.

Os atributos obrigatórios ocorrem sistematicamente em todas as descrições, exceptuando o atributo **endereço electrónico** das entidades responsabilidade e editor. Assim, ocorrem a 100% os seguintes atributos: **título próprio, descrição, língua,**

**assunto, tipo de recurso, URL, responsabilidade – nome, responsabilidade – função, editor – nome, editor – função.**

Os atributos com estatuto facultativo ocorrem com diferentes representatividades e são os seguintes, a saber: **título alternativo** que ocorre em 18%, **audiência** em 98%, **cobertura – geografia** em 18%, **cobertura – data início / data fim** em 14%; **cobertura – período de tempo** em 46%, **data de criação** (da informação) em 32%. Ao contrário destes casos, o atributo **relação** pode ter várias instanciações pelo que o cálculo foi feito tendo em conta todas as ocorrências, verificando-se assim a sua presença em 78% das descrições. De entre as várias instanciações possíveis a mais representativa é “Recurso requer” com 23 ocorrências, seguido de “É parte de” com 11 ocorrências, “Faz referência a” com 8 ocorrências, “É referido por” com 3 ocorrências e, finalmente, “Baseado na cópia de” com duas ocorrências. Globalmente verifica-se a presença, em média, de 3 (2.8) atributos facultativos em cada registo. A grande representatividade dos atributos Audiência, Cobertura – período de tempo e Relação acaba por lhes conferia um estatuto de importância semelhante ao dos campos obrigatórios, sobretudo para o primeiro caso, já que ele constitui uma das sub-categorias da estrutura de *browsing* dos recursos (Cf. Ponto 4.3.2)

O atributo **responsabilidade** está presente 96 vezes na amostra, sendo 84 ocorrências de autores pessoas físicas e 12 de autores pessoas colectivas, o que conduz a uma média simples de 2 (1.9) autores por descrição. A significativa ocorrência deste atributo, se bem que obrigatório, fica a dever-se também ao alargamento do conceito de autoria para funções como designer, programador, director, compilador e mesmo editor, certamente denotando o conceito de editor literário e não de editor da publicação, já que este é considerado uma entidade em si própria. Como já se referiu, dos atributos obrigatórios desta entidade nunca ocorre o Endereço Electrónico, pelo que seria eventualmente de excluir. Ao contrário, o único atributo facultativo – Afiliação – ocorre em 75% das referências à entidade, pelo que o seu estatuto se altera significativamente pela prática. Assim, é constante a ocorrência de atributos que permitem caracterizar de forma o menos ambígua possível esta entidade.

No que diz respeito ao atributo **título**, os **complementos de título** ocorrem em 42% das descrições e os **títulos alternativos** em 18%. Neste último caso, a maioria das ocorrências (5 em 9) são títulos abreviados, mais especificamente acrónimos, revelando-se desta forma uma prática que privilegia a forma completa como título próprio. Verificam-se apenas dois casos onde está prática não é seguida e são:

a) O título próprio *AAR syllabi Project* tem como título alternativo *American Academy of Religion syllabi Project*

b) Na descrição título próprio *ABTAPL union list of periodicals* não se desenvolve o acrónimo como título alternativo, sendo apenas descodificado em texto livre no atributo Descrição.

Os casos restantes configuram situações de títulos alternativos com conteúdo que ultrapassa a mera diferença formal ou linguística, como é o caso do título próprio *Contemporary postcolonial and postimperial literature in english* que recebe como título alternativo *Postcolonial and postimperial literature: an overview*. Nunca se verifica a ocorrência de títulos alternativos que são variantes do título próprio em outras línguas.

O atributo **URL** ocorre, como já referimos, em 100% das descrições e um teste sistemático à validade destas ligações revelou que apenas 3 delas se encontram desactivadas e em 5 verificou-se uma mudança de endereço mas com reencaminhamento do próprio recurso. Apenas em um caso se verifica a existência de um URL alternativo. Verifica-se uma completa ausência de outros indicadores.

Dos atributos que permitem concretizar os objectivos de identificação e selecção de um recurso e para além dos que já verificámos serem obrigatórios e ocorrerem a uma percentagem de 100%, e da percentagem dos que têm o estatuto de facultativos, verifica-se uma ausência do atributo **formato**, **data de actualização** (do recurso) e **versão**. No caso do atributo **requisitos de acesso**, a ocorrência significativa, já indicada, da instanciação “Recurso requer” do atributo Relação destina-se exactamente a indicar os requisitos, sobretudo tecnológicos, de acesso. A análise mais específica do atributo Assunto será realizada no ponto 4.3.1.

A prática de atribuição do atributo **relação** não contempla relações entre as descrições embora existam descrições relacionadas no catálogo como o demonstram os exemplos seguintes:

a) Exemplo 1: Na descrição de *Batkhin Centre home page* é atribuído o atributo Faz referência a com o conteúdo *Arts and Humanities Research Board*; embora este recurso também seja descrito no catálogo não existe qualquer ligação.

Exemplo 2: Na descrição do recurso *Creating and using virtual reality* é atribuído o atributo É parte de com o conteúdo *Arts and Humanities Data Service*, apesar de este recurso estar descrito no catálogo não existe qualquer ligação.

Grande percentagem das ligações relacionais referenciadas conduzem ao resultado esperado, exceptuando: duas ligações quebradas (sendo uma um endereço alternativo), e duas alterações de endereços não identificadas.

O atributo **assunto** é o único a partir do qual é possível navegar de uma descrição individual e aceder a outras descrições com o mesmo critério (Cf. Ponto 4.3.1).

No que concerne ao controlo de autoridade Autor, como já se verificou (Cf. Ponto 3.3.1) não está disponível a pesquisa por campos específicos, por isso não é possível testar inequivocamente. Não obstante e dado que não se verificam ligações dos atributos autor nas instanciações dos registos em que ocorrem para qualquer tipo de índice ou lista, podemos concluir com alguma probabilidade pela sua inexistência. Esta situação é, de certa forma, contraditória com as especificações na linguagem para a entidade responsabilidade, bem como com a prática descritiva seguida.

Na estrutura de *browsing* (ver ponto 4.3.2) verifica-se a concretização do objectivo localizar todos os recursos com o mesmo assunto. No entanto, como na pesquisa por palavra-chave não é possível pesquisar especificamente no campo assunto, não é possível verificar conclusivamente a existência de controlo de autoridade. Fica a dúvida se a possibilidade verificada no caso da estrutura de *browsing* não se deve apenas a um dispositivo de *design* do sítio. Uma outra característica concorre para a verificação deste objectivo que é a da utilização de vocabulário controlado que se verifica e será analisada com mais detalhe no ponto 4.3.1.

Verifica-se a inexistência da prática de Títulos uniformes, não existindo também campos em branco nem erros de formatação da informação nem de ortografia. Os acrónimos são sistematicamente desenvolvidos, como vimos, de forma exemplar, para o caso dos títulos.

A prática descritiva permite assim a concretização suficiente dos objectivos Encontrar uma entidade singular, Acessibilidade à entidade descrita e Identificação e Selecção do recurso apropriado à satisfação das necessidades do utilizador. Os objectivos localizar todos os documentos da mesma **Obra**, todos os documentos do mesmo autor e todos os documentos da mesma versão não são claramente atingidos. No caso do objectivo encontrar todos os recursos sobre o mesmo assunto, concluiremos sobre a sua prossecução com mais detalhe no ponto seguinte.

A amostra parece demonstrar que a prática aponta para uma maior exaustividade na descrição do que as Regras de Catalogação deste serviço fariam supor.

### **4.3. Análise e representação do conteúdo**

#### **4.3.1. Utilização das linguagens documentais para efeitos de Classificação e Indexação**

A grande maioria dos serviços utiliza linguagens documentais já estabelecidas e com fortes raízes no ambiente das bibliotecas. Em termos de classificações generalistas são utilizadas as três com maior representatividade mundial: a Classificação Decimal Universal (CDU), (SOSIG) a Classificação Decimal de Dewey (CDD) (PSIGATE, AGRIFOR, NATURAL), a classificação da Biblioteca do Congresso dos E.U.A (LCC) (VETGATE). São utilizadas igualmente as seguintes classificações especializadas: a classificação da Biblioteca Nacional de Medicina, também dos E.U.A (NLM) (OMNI, BIORESEARCH), a *ACM Computing Classification* (EEVL) e a *Mathematics Subject Classification* (MSC.200) (EEVL). O único serviço que não utiliza uma classificação já existente é HUMBUL.

A classificação NLM<sup>65</sup> teve a sua origem num relatório sobre a Biblioteca médica do exército norte-americano publicado em 1949, que recomendava o estabelecimento de uma nova classificação baseada nos princípios da LCC. A primeira edição desta classificação foi publicada em 1951.

A NLM cobre o campo da medicina e ciências relacionadas, utilizando as tabelas QS-QZ e W-WZ da LCC, por sua vez retiradas permanentemente desta última. Para assuntos de alguma forma relacionados com as temáticas básicas e para os materiais de carácter muito geral, a NLM pode ser complementada pela LCC, exceptuando as tabelas QM (*Human Anatomy*), R (*Medicine*) e QR (*Microbiology*) na medida em que se sobrepõem às já consideradas na NLM. As tabelas são compiladas separadamente e podem ser utilizadas independentemente. As notações seguem um princípio alfa-numérico, sendo o primeiro e segundo níveis alfabéticos e todos os outros numéricos. A

---

<sup>65</sup> Disponível em <http://wwwcf.nlm.gov/class>

este último nível, os códigos numéricos são utilizados de uma forma mais enumerativa do que hierárquica.

Os descritores das classes principais são dados de forma generalista e devem ser interpretados como incluindo para cada assunto o sistema fisiológico respectivo, a especialidade ou especialidades médicas relacionadas, as regiões do corpo mais directamente abrangidas e os campos do saber directamente relacionados. Dentro de cada classe, o primeiro nível de subdivisão é feito com base nos órgãos do corpo humano.

A classificação da Biblioteca do Congresso dos EUA foi criada para servir, em primeira instância, as necessidades de uma vasta biblioteca generalista com ambições universalistas que é a Biblioteca do Congresso.

A LCC divide o conhecimento humano em 21 classes principais, subdivididas hierarquicamente de forma generalista. As subdivisões dentro de cada classe obedecem mais a um princípio enumerativo do que hierárquico, o que conduz a muita repetição de detalhes. Cada classe é compilada separadamente e pode ser utilizada independentemente.

A notação é construída através de um sistema alfanumérico, onde os dois primeiros níveis (a classe principal e a primeira subclasse) são alfabéticos e os seguintes numéricos, até ao máximo de 4 dígitos.

Nesta caracterização geral da LCC importa destacar ainda os seguintes aspectos quanto à estruturação das classes: verifica-se uma tendência para ordenar as facetas relativas à forma e ao assunto antes das obras gerais sobre um tópico; a subdivisão alfabética dentro de cada classe é frequentemente utilizada, o que permite grande flexibilidade dos qualificadores individuais, mas exige controlo quanto ao cruzamento entre as notações; cada classe é desenvolvida por um processo de síntese de nível muito variável e só tem aplicação nessa mesma classe. Estes pontos fracos da LCC são eventualmente compensados com aquilo que parece ser a sua principal vantagem e que se traduz no facto de as classes serem criadas e desenvolvidas por especialistas nos temas que pretendem representar, apresentando por isso uma adequação óptima aos conteúdos dos recursos (ROWLEY:2000).

A Classificação Decimal de Dewey e a Classificação Decimal Universal partilham várias características comuns, a saber: universalidade na representação do saber, estrutura de base hierárquica, notação numérica que identifica cada um dos assuntos, desenvolvida de acordo com uma base decimal. Aliás, a CDU foi criada a partir da CDD.

A CDD foi criada por Melvil Dewey no final do século XIX (a primeira edição é de 1876), com o objectivo principal de encontrar uma forma de organizar as colecções de documentos em bibliotecas generalistas. Desde a sua fundação, a CDD tem vindo a ser constantemente actualizada e desenvolvida, para o que muito tem contribuído o facto de a gestão deste instrumento de organização da informação estar a cargo de uma agência bibliográfica nacional, a Biblioteca do Congresso dos E.U.A. (Introduction to Dewey decimal classification)

O seu carácter universalista traduz-se na representação de todos os domínios do conhecimento humano divididos em 10 classes. Cada uma das classes é de seguida subdividida hierarquicamente de acordo com uma lógica que vai do geral ao particular, o que significa que, exceptuando as classes principais, um determinado assunto é sempre uma parte subordinada de um outro mais geral.

O princípio da divisão hierárquica é também expresso pela utilização de notações em números árabes desenvolvidas numa base decimal, ou seja, cada uma das classes é subdividida em 10 outras e cada uma dessas pode ser subdividida em mais 10 e assim sucessivamente, um ponto decimal é normalmente introduzido a partir do terceiro dígito, permitindo um novo desenvolvimento na mesma base decimal. A extensão da notação traduz a posição hierárquica de um conceito. Um número a um determinado nível está normalmente subordinado a uma classe cuja notação é um dígito mais pequena; coordena-se com outra que seja representada pela mesma extensão de dígitos e é anterior a uma classe com um ou mais dígitos do que ela própria. Exemplificando:

600 – *Technology*

630 *Agriculture and related technologies*

636 – *Animal husbandry*

636.7 – *Dogs*

636.8 – *Cats*

Os conceitos de *Dogs* e *Cats* são mais específicos (ou seja, subordinados) a *Animal husbandry*, são igualmente específicos entre si e *Animal husbandry* é menos específico do que *Dogs* e *Cats*.

Para além da característica hierárquica da sua estrutura, a CDD é um esquema de classificação muito enumerativo, ou seja, pretende atribuir uma designação e respectiva notação para cada conceito entendido como válido (TAYLOR: 1999). Esta tendência estava mais presente nos primeiros tempos de desenvolvimento, onde praticamente não existiam formas de construção de notações para além das que estavam presentes nas tabelas principais. Mais recentemente, a CDD tem evoluído no sentido dos esquemas facetados através de processos de construção de notações que recorrem a tabelas auxiliares e a desenvolvimentos aconselhados pelos editores no contexto de cada classe. Este desenvolvimento no sentido da utilização de facetadas apresenta algumas ambiguidades, na medida em que a consistência na sua estrutura é subordinada à notação, porque na CDD o número é tudo (ROWLEY:2000). A subordinação à lógica do desenvolvimento numérico constitui ao mesmo tempo o ponto forte e o ponto fraco desta classificação, na medida em que é a lógica da notação que dita a ordem dos assuntos e não a aplicação consistente de um determinado critério de subdivisão.

A Classificação Decimal Universal foi criada por Paul Otlet e Henri LaFontaine, dois advogados belgas, a partir de 1885 e publicada pela primeira vez entre 1904 e 1907. Actualmente, é gerida por um consórcio internacional e é actualizada periodicamente.(About Universal Decimal Classification and the UDC Consortium)

Como já referimos a CDU foi criada com base na CDD, partilhando por isso com esta as características básicas. Dois importantes aspectos as distinguem: em primeiro lugar, desde o início que a CDU desenvolveu de forma mais específica a grande maioria dos assuntos em todas as tabelas, o que não raro produz notações de uma extensão

considerável; em segundo lugar, a possibilidade de representar assuntos mais complexos e identificar com clareza o ponto de vista sob o qual é abordado um determinado conceito é mais eficaz, na medida em que as notações retiradas das tabelas principais podem ser acompanhadas por outras retiradas de várias tabelas auxiliares e que permitem exprimir uma ou várias facetas do assunto. Existem tabelas auxiliares para atribuir facetas de forma (por exemplo, a tabela Id de auxiliares comuns de forma permite particularizar a forma de apresentação de um dado conteúdo), de tempo, de lugar, de língua, etc.

Para além das tabelas auxiliares gerais, dentro de algumas classes existem desenvolvimentos específicos que se traduzem pela existência de uma espécie de tabelas auxiliares para essa classe, denominadas subdivisões auxiliares especiais, podendo uma dessas subdivisões ser utilizada com determinadas notações principais dessa classe. Exemplifiquemos: para indicar que uma Obra de literatura portuguesa pertence ao género do romance policial deve encontrar-se a notação correspondente para literatura portuguesa na classe 8, neste caso, 821.134.3 adicionando-lhe em seguida a notação de um das subdivisões auxiliares especiais da classe 8, a dos géneros literários, ou seja -312.4 (o hífen é um dos sinais que distingue a notação principal de uma notação auxiliar). Assim, teríamos como notação neste exemplo: 821.134.3-312.4. A CDU aposta, desta forma, menos na enumeração de todos os conceitos, do que na construção de notações que permitem traduzir de uma forma mais flexível a diversidade de abordagens ou pontos de vista sob o qual um assunto ou conceito pode ser tratado.

As diferenças entre a CDU e a CDD estão intimamente ligadas à forma como estes instrumentos foram encarados pelas comunidades biblioteconómicas dos E.U.A e da Europa. Enquanto que para os primeiros, as classificações foram encaradas mais como uma forma de organizar fisicamente os documentos na estante, pelo que notações complexas com vários auxiliares não são adequadas; para os segundos, elas foram sempre mais perspectivadas como uma forma de representar o conteúdo dos documentos para efeitos de recuperação no contexto dos catálogos, pelo que se exigia uma maior especificidade e flexibilidade.

A MSC 2000 (*Mathematics subject classification*)<sup>66</sup> organiza a representação do conhecimento sobre esta temática em 97 classes (nem todas actualmente preenchidas), atribuindo a cada assunto uma notação alfa-numérica. O desenvolvimento de cada tema é realizado, aparentemente, a um nível de especificidade máximo relativamente superficial já que apenas se consideram 2 subdivisões da classe principal. Esta opção produz o extenso número de classes principais já referido, o que permite contrabalançar o reduzido desenvolvimento de cada classe. Assim, as classes principais apresentam desde logo uma certa especificidade de representação. Por exemplo, o que poderíamos considerar como uma das grandes classes principais – a Álgebra – está representado, pelo menos, através de 14 classes principais. As notações têm uma carácter mais enumerativo que hierárquico. As classes relacionam-se entre si pela técnica das referências de exclusão (VER) e de explicação (VER TAMBÉM) todas explicitamente definidas na linguagem. A aplicação desta linguagem distingue claramente classificações principais de classificações secundárias, sendo estas últimas consideradas referências cruzadas. Aconselha-se, portanto, claramente a atribuição tendencial de uma notação a cada documento classificado.

---

<sup>66</sup> Disponível em <http://www.ams.org/msc/classification.pdf>

A *ACM Computing Classification System* (1998)<sup>67</sup> desenvolve-se em 11 grandes temas que pretendem cobrir todos os aspectos do campo científico da computação, desde *Hardware, Software*, até *Teory of Computing, Computing Mehodologies* para terminar em *Computing Milieux*. Estes nós principais subdividem-se em dois ou três níveis, começando todos pelo assunto *General* e terminando em *Miscellaneous*. O primeiro nível é representado em termos notacionais por letras do alfabeto (de A a K), e os segundos e terceiros níveis por notação alfa-numéricas.

Os primeiros níveis da classificação não são normalmente utilizados para classificar os materiais informativos, para classificar informação de carácter generalista utiliza-se o primeiro sub nó de cada assunto geral, (por exemplo C.0 *Computer systems General*).

Associados a cada nível podem encontrar-se descritores de assunto não notados. Pretendia-se com esta estratégia uma fácil alteração destas representações no sentido de adaptar facilmente a classificação à evolução dos conhecimentos. No entanto, a dificuldade em alterar os descritores considerando que durante a sua vigência centenas ou milhares de documentos foram classificados fazendo recurso a estes, transformaram-nos em parte integrante das tabelas.

Para além destes descritores de assunto, é possível associar apropriadamente em cada nível descritores implícitos que se traduzem na possibilidade de utilizar nomes próprios para especificar o assunto (por exemplo, C++ é um descritor implícito de D.3.2. *Language classifications*).

Esta linguagem apresenta-se, assim, como uma classificação com um reduzido desenvolvimento específico das classes principais, o que é lógico tendo em conta a especificidade temática de origem. Os descritores não notados, embora semanticamente dentro da estrutura, apresentam-se um pouco fora da lógica das notações e tendo em conta a sua tendencial permanência fica difícil explicar o porquê desta especificidade.

Quatro outras linguagens de tipo diferente são utilizadas pelos serviços em análise: *Medical Subject Headings* (MeSH) (BIORESEACRH, OMNI), o tesouro *CAB: the thesaurus for the applied life sciences*. (AGRIFOR, VETGATE), o tesouro *HASSET: Humanities and Social Science Electronic Thesaurus* (SOSIG) e o tesouros *Ei: Engineering Index* (EEVL). HUMBUL não utiliza qualquer linguagem de indexação e PSIGATE utiliza palavras-chave retiradas da linguagem natural, portanto vocabulário não controlado.

As MeSH<sup>68</sup> são constituídas por um conjunto de descritores que representam conceitos na área da Medicina. Esta linguagem é desenvolvida pela *National Library of Medicine* dos E.U.A, desde o final do século XIX, e é actualmente muito utilizada em todo o mundo por instituições ligadas à Medicina.

A lista dos descritores é composta por 21.973 termos, ao que se adicionam centenas de referências cruzadas, totalizando milhares de pontos de acesso a todo o conjunto.<sup>69</sup> Para

---

<sup>67</sup> Disponível em [http://www.informatik.uni-stuttgart/zd/buecherei/fibib\\_hilfe\\_cr.html](http://www.informatik.uni-stuttgart/zd/buecherei/fibib_hilfe_cr.html)

<sup>68</sup> Disponível em <http://www.nlm.nih.gov/mesh/filelist.html>

<sup>69</sup> Os dados quantitativos aqui referenciados foram retirados de "Fact Sheet Medical Subject Headings".

além de uma lista alfabética, a terminologia é organizada hierarquicamente, através de estruturas em árvore (15 estruturas subdivididas em várias subcategorias), como se pode verificar pelo exemplo que se segue:

Apresentação estrutural dos termos de MeSH para o termo *Anatomy*

- . *Anatomy*
  - . *Body regions*
    - . *Abdomen*
      - . *Abdominal cavity*
      - . *Abdominal wall*
      - . *Groin*
      - . *Inguinal canal*
      - . *Umbilicus*
    - . *Back*  
(etc.)

Para além dos descritores principais, estão também disponíveis 82 descritores secundários ou qualificadores, que podem ser usados para especificar melhor um conceito que se pretende representar. Os descritores secundários permitem representar aspectos de abordagem que são comuns a muitas das temáticas representadas pelos descritores principais, por exemplo: diagnóstico, educação, história, patologia, terapia, anatomia e histologia.

Algumas especificidades desta linguagem podem tornar-se um obstáculo para utilizadores menos experientes: a utilização preferencial de termos clínicos em vez de termos mais comuns (por exemplo, o uso de neoplasmas em vez de cancro ou tumores) apesar das referências de termos não autorizados para termos autorizados; o vocabulário é seleccionado a partir de várias línguas, o que corresponde à origem da terminologia médica dos E.U.A; a possibilidade de inverter a ordem dos termos, apesar de ter a vantagem de manter a sequência alfabética; a distinção entre sistemas e órgãos do corpo e as suas doenças, e finalmente o facto de alguns descritores principais poderem ser também utilizados como descritores secundários.

O tesouro CAB: *the thesaurus for the applied sciences* tem a sua origem numa organização internacional não lucrativa com a designação de CAB International, cujo objectivo é desenvolver a investigação e encontrar soluções para os problemas mundiais na área da Agricultura e das Ciências do Ambiente. Uma forte componente desta instituição é a publicação de documentação e a constituição de bases de dados de referências sobre as temáticas referidas. O tesouro CAB surge assim, e em primeiro lugar, para satisfazer as necessidades deste organismo em estruturar e disponibilizar acesso à informação por si produzida.

O CAB é actualmente constituído por 59.000 termos, que pretendem representar as seguintes áreas temáticas principais: Agricultura, Horticultura, Protecção das Colheitas, Florestas, Ciência dos Solos, Produção Animal, Veterinária, Nutrição Humana, Estudos Rurais, Entomologia, Parasitologia e Micologia. As relações entre os termos obedecem à tipologia comum, ou seja, são implementados os seguintes tipos de relações: Use, Usado por, Termo específico, Termo geral, Termo relacionado. Existem igualmente notas explicativas para muitos termos que esclarecem o sentido da sua utilização.

O Tesouro HASSET<sup>70</sup> é baseado no tesouro da UNESCO e desenvolvido em estreita cooperação com *The Data Archive*, espelhando as 5.000 áreas temáticas presentes neste serviço. (MILLER:1997). São implementados seis tipos básicos de relações entre termos neste tesouro: Use, Usado por, Termo específico, Termo Geral, Termo de topo e Termo relacionado. Para além disto existem também Notas explicativas para um grande número de termos que esclarecem o sentido da sua utilização. Em 1997 existiam aproximadamente 8.650 termos neste tesouro, 2.500 dos quais são termos não preferenciais ou sinónimos. As relações entre os termos atingem o valor de 38.600 formando 296 hierarquias. (MILLER:1997).

O *Gateway SOSIG* utiliza, para além deste tesouro, dois outros construídos localmente, a saber: *Social Work and Welfare Thesaurus* derivado de *Care Data*<sup>71</sup>, uma base de dados de publicações periódicos principalmente na área do Serviço Social, que implementa uma lista de cabeçalhos; e o *Government, Politics and Anthropology Thesaurus*, derivado da lista de descritores desenvolvida pela *IBSS*<sup>72</sup> (*International Bibliography of the Social Sciences*).

Considerando as temáticas abrangidas pelos serviços (cf. Ponto 3.3), verifica-se que a escolha das linguagens de classificação e indexação é em todos os casos particularmente apropriada. A opção por classificações de âmbito generalista apresenta a vantagem de utilizar a mesma linguagem que disponibiliza a possibilidade de representar um grande leque de assuntos. A desvantagem desta opção reside na posição subordinada que, em alguns casos, assuntos centrais para o serviço ocupam no esquema classificatório, desvantagem que pode ser ultrapassada de várias formas, por exemplo, pela supressão dos níveis mais gerais tornando os níveis específicos mais visíveis; ou ainda a falta de especificidade na representação de algumas temáticas. A opção contrária, protagonizada neste contexto por EEVL, tem a vantagem de uma adaptação otimizada ao tema a tratar, mas a desvantagem, em termos de produção da informação de implicar o domínio de linguagens diferentes e em termos de utilizador o conhecimento também de diferentes concretizações da representação do conteúdo. A utilização de diferentes classificações, por outro lado, implica também a necessidade de operacionalizar a interoperabilidade entre os termos, o que acrescenta complexidade organizativa.

Do ponto de vista do utilizador final, podemos igualmente considerar que a selecção dos serviços é adequada, já que se trata de público com competências para a decodificação das linguagens, bem como habituado à terminologia específica utilizada para indexar os recursos.

As classificações e os tesouros foram seleccionados para desempenhar um papel diferenciado. As classificações são utilizadas, exceptuando o caso de EEVL e SOSIG, para representar o conteúdo a um nível de grande generalidade, donde a média de classes atribuídas por cada serviço aos recursos seja relativamente reduzida, a saber: HUMBUL com 2.2, PSIGATE com 1.3. No caso de BIOME, embora as descrições não permitam visualizar as notações atribuídas, a análise da estrutura de *browsing* permitiu verificar baixas taxas de atribuição para todos os serviços, exceptuando NATURAL, pelo que se pode concluir também por uma utilização de acordo com este paradigma. No caso dos serviços EEVL e SOSIG, a aposta parece ser a utilização das classificações

---

<sup>70</sup> Disponível em <http://www.data-archive.ac.uk/search/hassetsearch.asp>

<sup>71</sup> Disponível em <http://195.195.162.66/elsc/caredata/caredatasearch.htm>

<sup>72</sup> Disponível em <http://www.lse.ac.uk/collection/ibss>

para uma representação mais exhaustiva do conteúdo já que o número de classes em média atribuídas a cada recurso é claramente superior aos outros casos, a saber: SOSIG com 4.9 e EEVL com 4.2. Os tesouros são utilizados para permitir uma representação de grande especificidade e exaustividade, apresentando todos os serviços médias elevadas de descritores atribuídos, com o valor extremo para EEVL de 15.8, mas na sua grande maioria duplicando o número de notações, como se verifica: PSIGATE com 6.4, SOSIG com 6.7 e BIOME com 3.5.

Assim, pode concluir-se que a prática de representação do conteúdo aponta para a utilização de duas diferentes linguagens usadas para implementar abordagens mais generalistas e mais exaustivas a essa representação.

A aplicação destas linguagens é feita de forma completamente atomizada, quer dizer, não são implementados quaisquer processos de relação entre elas, eventualmente através de um mapeamento terminológico que permitisse ao utilizador vias alternativas de resposta às suas questões.

Em todos os serviços, quer no caso das classificações quer nos dos tesouros, o seu emprego é feito de forma consistente. O baixíssimo coeficiente de variação (entre 0.4 e 1.6) face à média em todos eles confirma que a atribuição de notações e descritores é realizada ao longo de todos as descrições de acordo com a mesma perspectiva de utilização.

A partir de uma descrição o utilizador só tem acesso à relação entre os termos dos tesouros se clicar sobre um termo seleccionado, sendo que esta possibilidade só se verifica no conjunto de serviços do *hub* BIOME, podendo aí ter acesso às relações verticais e horizontais próprias deste tipo de linguagem. No caso das classificações, esta possibilidade nunca está presente.

A utilização das classificações não inclui qualquer referência à versão da linguagem a ser utilizada. Nunca se utilizam também as notações das classificações, apenas os descritores a elas associados.

Todas as linguagens prevêem o atributo categoria / tipo de recurso, definindo várias categorias de recursos. Analisando essas definições, verifica-se uma grande proximidade conceptual entre os vários serviços. São definidas categorias de acordo com vários critérios: formato, forma de apresentação, origem da informação, objectivos da informação. Assim, temos como categorias praticamente transversais as seguintes: Organizações, Livros, Publicações periódicas, Materiais áudio-visuais e multimédia, Publicações Governamentais, Artigos, Relatórios, FAQ's, Listas de Discussão, Bases de Dados (incluindo as bibliográficas), Projectos e Centros de Investigação, Software, Serviços de Notícias (News). Como se pode verificar, este atributo pode desempenhar importantes funções no que diz respeito à concretização dos objectivos de Identificação e Selecção, tendo em conta a sua proximidade semântica com a noção de representação do conteúdo.

### 4.3.2. Análise das estruturas de *browsing*

Todos os serviços analisados disponibilizam uma opção de *browsing* da colecção de recursos disponível, recorrendo para tal às linguagens documentais analisadas. Utilizando ora uma estrutura em árvore, para as classificações, ora uma lista alfabética, para os tesouros, estes serviços proporcionam uma forma alternativa de acesso à colecção para além da tradicional pesquisa por palavra-chave.

#### 4.3.2.1. SOSIG

Este serviço disponibiliza uma estrutura em árvore, a partir da página principal, com vários níveis de especificidade, construída com base na Classificação Decimal Universal.

Para uma vasta área de assuntos como aquela que este serviço pretende abarcar, a selecção da CDU para este efeito parece apropriada na medida em que permite categorizar os recursos seleccionados de acordo com um vasto conjunto de áreas temáticas, podendo essa categorização ser realizada a diferentes níveis de especificidade, o que pode ser particularmente adequado tendo em conta a variedade de tipologias de recursos que são consideradas.

O primeiro nível da estrutura traduz-se na selecção de 17 assuntos, cuja representação em termos de especificidade se situa, na grande maioria dos casos, ao nível de dois dígitos da Classificação Decimal Universal, o que significa ter-se optado por uma representação de tipo generalista já que este nível representado por dois dígitos é o que se situa imediatamente abaixo da definição das grandes classes. Em termos de CDU, as grandes classes aqui representadas são apenas a Classe 1 (Filosofia / Psicologia) e a Classe 3 (Ciências Sociais).

Os assuntos principais são em seguida subdivididos em outros níveis de maior especificidade até ao limite de 4 níveis (Tabela 23).

Verifica-se uma clara predominância do segundo nível de especificidade, o que traduz uma reduzida exploração das capacidades da Classificação Decimal Universal na representação de um qualquer assunto. A grande diferença entre o primeiro e o segundo sub-níveis resulta do maior desenvolvimento dado ao assunto Direito (*Law*) ao ter-se optado por uma subdivisão no segundo nível de acordo com um critério geográfico específico, ao nível do país.

Em termos globais, verifica-se, de qualquer forma, que o número de subdivisões em cada classe é significativo, oscilando entre um mínimo de 4 e um máximo 411, ou se excluirmos o caso particular do assunto *Law*, entre 4 e 127. Esta amplitude pode ter origem em diversas circunstâncias, a saber: a diferença entre os editores responsáveis por cada uma das secções, o momento mais precoce da entrada do tema na estrutura e a centralidade do tema em termos da cobertura de assuntos definida para o serviço.

Ainda ao primeiro nível da árvore é possível fazer actuar uma restrição de tipo geográfico em três categorias diferentes: *World*, *Europe*, *UK*. O resultado de qualquer

destas restrições é a limitação ou alargamento do conjunto de recursos disponíveis em cada assunto.

Após a selecção de um assunto, os recursos encontram-se organizados de acordo com a seguinte tipologia dividida por categorias de serviços (a saber: *Articles/Papers/Reports (collections)*, *Bibliographic Databases*, *Companies*, *Datasets*, *Educational Materials*, *Governmental Bodies*, *Journals (contents and abstracts)*, *Journals (full text)*, *Mailing Lists/Discussion Groups*, *News*, *Organisations/Societies*, *Research Projects/Centres* e *Resource Guides*) e categorias de documentos (a saber: *Articles/Papers/Reports (individual)*, *Bibliographies*, *Books/Book Equivalents*, *Faq's*, *Government Publications*, *Reference Materials*).

É possível navegar no conjunto de categorias temáticas através da presença de uma tabela onde se encontram representadas. Uma hiperligação conduz o utilizador directamente à categoria seleccionada. Ainda neste nível, ou seja, após a selecção de um determinado tema que se encontre no sub-nível 2, o utilizador obtém informação sobre as secções relacionadas no catálogo, às quais pode aceder através da respectiva hiperligação.

Em termos de possibilidade de navegação nos assuntos, pode navegar-se na subdivisão hierárquica a partir da página principal, que apresenta o assunto geral e maioritariamente dois dos sub-assuntos de cada categoria. A partir da página principal é também possível aceder a uma lista completa dos assuntos que pode ser visualizada de duas formas: pela estrutura hierárquica dos assuntos e alfabeticamente para todos os assuntos independente do nível na hierarquia. Por defeito, a lista encontra-se estruturada pelos níveis hierárquicos de cada assunto.

Se o utilizador optar, a partir da página principal, por seleccionar um assunto de nível geral, são-lhe mostradas em seguidas as subdivisões de segundo nível. O sinal de + (mais) à frente de cada uma dessas subdivisões indica a existência de um outro nível. Em cada uma das subdivisões é sempre possível aceder a uma lista de assuntos da categoria geral onde essa subdivisão se integra.

A possibilidade de atribuir o mesmo recurso a mais do que um assunto é claramente explorada apresentando uma taxa de 103.6%.

Os recursos são sobretudo classificados (Cf..Tabela 24) nos segundos (46.7%) e primeiros níveis (41.6%), o que corresponde também à existência de um maior número de assuntos. Apenas em um caso não existem recursos classificados nos temas gerais, o que parece revelar uma opção por classificar a um nível de maior generalidade.

O número de recursos atribuídos em cada assunto revela um nítido privilégio para algumas temáticas, nomeadamente *Politics* (14%), *Economics* (13.3%), *Law e Government and Public Administration* (11.5%). Verifica-se assim que menos de 30% dos assuntos recolhe praticamente 51% dos recursos. Esta desigual distribuição explica-se por várias ordens de razões: a cooperação com instituições especializadas nessas temáticas; o momento, no tempo, em que o tema começou a ser objecto de representação; e finalmente a própria natureza do assunto, na medida em que nem todos os temas seleccionados como assuntos principais estão ao mesmo nível de generalidade, pelo que seria impossível, por exemplo, que temas como *Social Sciences General* ou

*Women Studies* atingissem o mesmo número de recursos que temas como o *Law* ou *Economics*.

O número de recursos atribuído em cada assunto é relativamente equilibrado (Cf. Tabela 25). Os valores excessivos dos primeiros dois intervalos (1-10 recursos e 11-20 recursos) resultam sobretudo do desenvolvimento atípico do tema *Law*. De realçar a importância dos intervalos acima dos 100 recursos, que representam 16.3% de todos os recursos atribuídos. A desvantagem de ter que percorrer sucessivas páginas de referências para analisar a pertinência de alguma delas para resposta a uma necessidade é compensada pela possibilidade de pesquisa por palavra quer ao nível do tema geral, quer ao nível da secção específica em que o utilizador se encontra.

Daqui decorre também que o número de assuntos com poucos recursos classificados (até 5) não é significativo (Cf. Tabela 26), mesmo considerando o caso especial do Direito já referido.

#### **4.3.2.2. HUMBUL**

Este serviço apresenta não exactamente uma árvore, mas um conjunto de assuntos de carácter generalista, disponíveis a partir da página principal:

##### **Lista de Assuntos em HUMBUL:**

*Archaeology*

*History*

*History and Philosophy of Science*

*American Studies*

*Linguistics*

*German Studies*

*Spanish Studies*

*Russian and Slavonic Studies*

*Classics*

*Philosophy*

*Religion and Theology*

*Humanities computing*

*English Studies*

*French studies*

*Italian Studies*

*Comparative literature*

Se comparada com a abrangência temática do serviço, este conjunto de assuntos é particularmente pobre. Como se verifica, o nível de generalidade é muito alto, pretendendo cobrir apenas as grandes temáticas que são objecto do serviço. Não existe nenhum sub-nível temático de divisão. Este conjunto de assuntos tem origem na definição das temáticas abrangidas pelo serviço, sendo utilizada exactamente a mesma lista de assuntos. Embora se refira que no primeiro trimestre de 2001 será aplicada uma classificação à organização dos recursos, chegando a sugerir-se a LCSH<sup>73</sup>, tal não se verificou ainda. Neste âmbito ainda, a sugestão de que este conjunto generalista de

---

<sup>73</sup> Cf. Fraser, Michael; Edwards, Alun – [Describing and cataloguing resources](http://www.humbul.ac.uk/about/catalogue.html#subject). Disponível em <http://www.humbul.ac.uk/about/catalogue.html#subject>

assuntos poderá vir a ser desenvolvido de acordo com a tabela de subdivisão de assuntos utilizada pela HESA (*Higher Education Statistics Agency*) também não chega a ser concretizada.<sup>74</sup>

Após entrar em cada um dos assuntos, o utilizador encontra apenas um conjunto de cabeçalhos de natureza formal que organizam os recursos disponíveis de acordo com as seguintes categorias:

**Tipos de Recursos:**

*Projects / Organizations*

*Primary sources*

*Research Related*

*Secondary sources*

*Teaching / Learning related*

*Bibliographic sources*

Esta categoria é a que se encontra definida por defeito. Para aceder às seguintes, o utilizador necessita de seleccionar a opção pretendida, a saber:

**Período de Tempo:**

3.000 – 1500 AC

1.500-750 AC

750-1 DC

1-500 DC

500-1000 DC

1000 – 1200 DC

1200-1500 DC

Século 16

Século 17

Século 18

Século 19

Século 20

No caso da Arqueologia são adicionados os seguintes períodos:

2.5 milhões – 150.000 AC (Paleolítico Inferior)

150.000-40.000 AC (Paleolítico Médio)

40.000-8.000 AC (Paleolítico Superior)

8.000-5.000 AC (Neolítico Inicial)

5.000-3.000 AC

**Audiência:**

*Postgraduate / Faculty*

*Undergraduate*

*General Public*

A selecção destes cabeçalhos formais parece ser particularmente adequada às temáticas tratadas pelo serviço, sobretudo as suas primeiras categorias. De facto, a distinção entre Fontes Primárias e Secundárias é particularmente relevante na área das Humanidades,

---

<sup>74</sup> Cf. Fraser, Michael; Edwards, Alun – Collection development policy. Disponível em <http://www.humbul.ac.uk/about/colldev.html>

onde é pertinente a distinção entre os “textos” históricos e toda a literatura produzida sobre eles ao longo dos tempos. Pelo mesmo tipo de motivos, a definição de uma tabela cronológica assume grande pertinência.

Finalmente, o utilizador pode ainda aceder a uma lista de todos os recursos atribuídos em cada tema, o que permite navegar na totalidade sem a atribuição múltipla que se verifica se seleccionadas qualquer uma das outras categorias.

De facto, se considerarmos a classificação de recursos pela categoria que é disponibilizada por defeito, verifica-se uma alta taxa de atribuição, na ordem dos 105%, visto que, embora sejam apenas classificados 4868 recursos, a sua contabilização em todas as instâncias da referida categoria apresenta o valor de 9969.

A distribuição dos recursos pelos temas permite avaliar a posição relativa destes no contexto do serviço (Cf. Tabela 27). Verifica-se uma predominância dos seguintes temas, por ordem de importância: *History*, *Archaeology*, *English Studies*, *Religion & Theology* e *Philosophy*.

Alguns assuntos apresentam valores residuais com menos de 1% de recursos classificados, a saber: *Spanish Studies*, *Italian Studies* e *Comparative Literature*.

#### **4.3.2.3. PSIGATE**

As áreas temáticas cobertas por este serviço são organizadas individualmente em estruturas hierárquicas de assuntos com base na Classificação Decimal de Dewey. Assim, não é possível, a partir de um ponto onde estejam representados todos os temas aceder à totalidade da colecção.

Em qualquer um dos temas, os assuntos principais são subdivididos em dois níveis de especificidade (Cf. Tabela 28).

Exceptuando o assunto *Sciences History / Policy*, todos os outros são desenvolvidos a dois níveis de especificidade, sendo no assunto *Physics* que tal subdivisão apresenta um maior desenvolvimento quantitativamente considerado.

A análise da estrutura de desenvolvimento de um dos temas permite verificar que nem sempre o nível em que estes se situam na estrutura de *browsing* definida coincide com a sua posição hierárquica no âmbito da Classificação Decimal de Dewey (Cf. Tabela 29)

Verifica-se que, quer no primeiro, quer no segundo nível do directório, os assuntos são seleccionados em diferentes níveis hierárquicos da Classificação Decimal de Dewey.

No primeiro nível, a utilização da CDD está ao nível dos 3 dígitos basicamente, o que significa um nível de especificidade pouco desenvolvido. Tal tendência só é ultrapassada nos casos em que a importância do tema é superior ao seu nível de representação na CDD: isto verifica-se em 33% dos assuntos seleccionados.

No segundo nível, a utilização da CDD desenvolve-se a partir do nível anterior, atingindo um nível de especificidade não superior ao que é possível representar através

de 3 dígitos. São apenas 4 os casos em que tal não se verifica. Por outro lado, o expediente de utilizar a mesma notação para designar assuntos considerados de forma autónoma permite exactamente resolver todas as situações em que a falta de especificidade de representação desses assuntos na tabela impediria a sua fácil individualização e esses assuntos assumem uma grande centralidade no contexto das temáticas do serviço.

A organização da estrutura dentro de cada um dos grandes assuntos segue um mesmo padrão: após a selecção de um dos temas, acede-se a uma lista ordenada alfabeticamente, que contém todos os sub-assuntos do primeiro nível. Seleccionando um desses sub-assuntos, é mostrada uma página com uma primeira lista de recursos e com informação sobre a posição que o termo seleccionado ocupa na hierarquia de desenvolvimento do assunto, ou seja, é indicado o termo que ocupa a posição hierárquica imediatamente superior e todos os termos subordinados. Desta forma, é sempre possível uma percepção do contexto em que se situa o termo a que se acede, pelo menos na sub-área temática seleccionada. De facto, não são estabelecidos quaisquer tipos de referências cruzadas entre os termos de uma mesma área de assunto, ou seja, nem entre as subdivisões de primeiro nível nem entre estas e as de segundo nível, nem mesmo entre os termos de um mesmo nível hierárquico que pertencem a áreas temáticas associadas.

Em nenhum dos níveis da estrutura é dada informação sobre as notações correspondentes a cada um dos termos seleccionados.

Em termos de funcionalidades de navegação na estrutura, o utilizador encontra sempre presente informação sobre o caminho que percorre nessa estrutura, com uma hiperligação que permite aceder directamente à página inicial e à grande área temática em que se encontra, mas estranhamente não ao nível imediatamente anterior se se encontrar num termo de segundo nível. Exemplifiquemos: quando o utilizador selecciona o sub-assunto *Asteroids* que faz parte do assunto *Solar System*, integrado no tema *Astronomy*, tem acesso à seguinte informação sobre o caminho percorrido - *Home* → *Astronomy* → *Asteroids* – onde se ignora o assunto imediatamente anterior a *Asteroids*, ou seja, *Solar System*. Para voltar a aceder ao termo *Solar System* tem que usar a funcionalidade de retroceder do *browser*.

Toda a informação sobre termos subordinados ou mais gerais contém hiperligações, permitindo assim que o utilizador possa aceder directamente a todos os níveis hierárquicos. O acesso a uma subdivisão de primeiro nível obriga, no entanto, à voltar a página inicial do tema, o que é disponibilizado através da já referida hiperligação sobre o percurso percorrido, visto que esta informação contém sempre uma hiperligação para o termo de topo, ou seja, o termo inicial.

A informação sobre o número de recursos disponíveis é apenas fornecida individualmente para cada uma das categorias de assunto, independentemente do nível hierárquico em que essa categoria se encontre posicionada. Ou seja, se uma determinada categoria se subdivide em outras categorias seria esperável que o utilizador encontrasse informação sobre o número de registos classificados apenas na categoria mais abrangente, todos os classificados em cada uma das categorias subordinadas, mas também o somatório de todos os recursos classificados. Exemplifiquemos: quando o utilizador selecciona o assunto *Solar System* (um dos assuntos do tema *Astronomy*)

encontra informação sobre o número de registos classificados nessa categoria (80), e se aceder a cada uma das sub-categorias de *Solar System* encontra também informação sobre o número de recursos em cada uma delas, mas nunca lhe é dado o número total de registos de *Solar System* e de todas as suas categorias subordinadas.

Apesar de o serviço conter apenas 7.100 descrições de recursos, o número de recursos contabilizados na estrutura de *browsing* ascende a 9.792, o que traduz uma taxa de atribuição de 38%.

Os recursos são sobretudo classificados no primeiro nível de especificidade dos assuntos, apesar de ser no segundo nível que se encontra o maior número de assuntos (Cf. Tabela 30). O número de recursos atribuídos ao primeiro nível representa 66.5%, enquanto que os atribuídos ao segundo apenas 33.4%. Só no caso do tema *Chemistry* são atribuídos mais recursos no segundo nível. Considerando que existem sempre recursos atribuídos ao primeiro nível mesmo que este se subdivide, a opção parece ser a de uma representação a um nível de maior generalidade.

A distribuição dos recursos por cada grande tema revela um nítido privilégio para três dos seis temas tratados: no primeiro nível os recursos atribuídos aos temas *Chemistry*, *Earth Sciences* e *Physics* representam 67% de todos os recursos atribuídos.

Um conjunto relativamente reduzido de recursos é atribuído em muitos dos assuntos (Cf. Tabela 31). A grande maioria dos assuntos apresenta um nível de atribuição de recursos entre 1 e 20 (45.1%).

Poucos assuntos contêm um número elevado de recursos atribuídos, de forma que o problema da necessidade de consultar várias páginas até seleccionar os que melhor se adequam à resposta a uma dada necessidade não é, nesta fase, muito emergente. De qualquer forma, este serviço não implementa formas de ultrapassar este problema num futuro próximo, como já o fazem EEVL ou SOSIG..

Se considerarmos que um assunto que contenha apenas um número até cinco recursos classificados é assunto escassamente representado, verifica-se que o volume desta ocorrência não é muito significativo (Cf. Tabela 32).

O conjunto dos assuntos com um volume até cinco recursos atribuídos representa apenas 12.9% da totalidade dos assuntos, o que poderá estar relacionado com o facto de os assuntos estarem apenas representados a dois níveis de especificidade.

A repetição do mesmo assunto em diferentes classes pode ser um resultado da aplicação das classificações na medida em que a sua divisão por áreas do saber implica muitas vezes que um mesmo assunto pode estar representado em classes diferentes dependendo do ponto de vista segundo o qual é analisado. Assim, por exemplo o assunto *Crianças* se for abordado do ponto de vista da Psicologia encontrará a sua representação na classe 1, mas se abordado do ponto de vista da Educação terá que ser representado na classe 3. Donde neste contexto, onde a ligação entre partes da informação está facilitada, a importância das referências cruzadas. Uma situação diferente desta, que configura alguma incongruência na utilização do vocabulário e na estruturação do *browsing*, é a da repetição do mesmo assunto, sob o mesmo ponto de vista, em diferentes classes. Esta situação verifica-se para os seguintes assuntos: *History of Astronomy*, *History of Chemistry*, *History of Earth Sciences* e *History of Physics*, que se encontram

representados todos na área de *Science / History Policy* e cada um deles na área temática específica que lhes corresponde, ou seja, *Astronomy, Chemistry, Earth Sciences* e *Physics*. Esta situação resulta provavelmente da adição posterior do tema *Science History / Policy* num momento em que os aspectos da história de cada ciência tinham já sido identificados aquando do desenvolvimento anterior de cada uma das ciências. De qualquer forma, a manutenção desta repetição é reveladora, estamos em crer, de alguma inconsistência na construção e desenvolvimento da estrutura de *browsing*, na medida em que o problema de eventualmente sinalizar ao utilizador a existência do tema História de... em cada uma das áreas temáticas poderia ser resolvido com uma referência para a área *Science History / Policy*.

#### 4.3.2.4. BIOME

Dado que não é disponibilizado nenhum mapeamento entre as diversas linguagens que os serviços que integram BIOME utilizam, nem entre aqueles que utilizam a mesma, nem qualquer outra forma que permita realizar o *browsing* em conjunto, cada um dos serviços apresenta um interface individual onde desenvolve autonomamente as suas opções.

##### AgriFor

Este serviço apresenta três formas diferenciadas de *browsing*<sup>75</sup>: um directório de assuntos que utiliza, ao que podemos verificar, uma linguagem não normalizada; uma lista de descritores retirados da CDD e uma lista de descritores retirada do tesauro CAB.

No primeiro caso, são seleccionados 10 assuntos, se bem que nem todos ao mesmo nível de generalidade, subdivididos, no máximo em três níveis de especificidade. Apresenta-se, assim, uma estrutura em árvore de base temática onde os assuntos seleccionados se organizam de acordo com um princípio de relação temática.

Verifica-se (Cf.Tabela 34) que em termos quantitativos, o número de sub-níveis é significativo, sobretudo se tivermos em conta que se trata de um tema já específico. Uma clara aposta é feita num nível intermédio de especificidade, se bem que o número de níveis existentes no sub-nível 3 não deixa de ser significativo.

No entanto, nem todos os temas conhecem o mesmo nível de especificidade de representação. Tal poderia eventualmente relacionar-se com a maior ou menor centralidade do tema relativamente às temáticas abrangidas pelo serviço. No entanto, este motivo não explica, por exemplo, porque razão um tema como *Economics, Trade & Rural Development* se encontra desenvolvido até ao 3º sub-nível e temas como *Soil Science e Fisheries* vão apenas até ao 2º sub-nível. Uma outra hipótese explicativa para esta diferença poderia estar relacionada com o nível de generalidade / especificidade do assunto principal: quanto mais geral for o assunto principal maior probabilidade existe de um maior número de subdivisões. De facto, se analisarmos os três assuntos que não atingem o sub-nível 3, verifica-se que são assuntos mais específicos à partida do que todos os outros.

---

<sup>75</sup> Disponível a partir de <http://agrifor.ac.uk/browse>

O nível de atribuição de recursos a vários assuntos é reduzido, apresentando uma taxa de 3.3%. Não existe aparentemente nenhuma razão específica para uma tão reduzida taxa de atribuição, a não ser uma menor importância dada a esta possibilidade pelos produtores da informação.

Os recursos nunca são atribuídos a um nível mais geral se este se subdivide em outros mais específicos e a um maior número de assuntos no segundo nível corresponde também o maior número de recursos atribuídos no mesmo nível: 51% dos recursos são aí classificados (Cf. Tabela 35).

A vasta maioria dos assuntos tem um reduzido número de recursos atribuídos: 72% dos assuntos tem apenas entre 1 e 10 recursos (Cf. Tabela 36). O reduzido valor de todos os intervalos, a partir inclusivé de 21-30, atesta claramente uma concentração à volta dos dois primeiros intervalos atingindo 329 assuntos, o que representa quase 90% dos assuntos presentes (88.4%).

Se considerarmos que assuntos com um número de recursos que oscilam entre 1 e 5 são assuntos escassamente representados e se contabilizarmos as ocorrências, verificamos a acentuação da tendência já detectada (Cf. Tabela 37). Estes assuntos com poucos recursos atribuídos representam uma importante fatia de todos os assuntos (41.6%).

Nem todos os temas conhecem o mesmo desenvolvimento quantitativo em termos do número de recursos atribuídos (Cf. Tabela 38). A balança pesa decisivamente para os seguintes temas, por ordem de importância: *Agriculture, Forestry, Economic & Trade and Rural Development, Food e Livestock*. Os outros assuntos apresentam valores muito abaixo destes, revelando uma *décalage* que acentua o pouco equilíbrio na representatividade dos temas.

O acesso através da CDD é disponibilizado a partir de uma lista alfabética de descritores retirados da tabela ao nível de especificidade representado pelos 3 dígitos da classificação (Cf. Tabela 39).

O número total de assuntos é de 63, representando a classe 6 da CDD o maior peso, como seria de esperar, dado que é nessa classe que se encontram representados a grande maioria dos assuntos das temáticas abrangidas por este serviço (Cf. Tabela 40).

A maior percentagem de notações retiradas da classe 3 aponta para uma preocupação de seleccionar aspectos ligados com a Economia e Sociologia Rurais, em detrimento de outros, por exemplo, os relativos à investigação científica pura, dados pela classe 5.

A utilização dos descritores retirados do tesouro CAB é disponibilizada através de uma lista alfabética, podendo o utilizador percorrê-la através das letras do alfabeto. O nível de especificidade na representação dos assuntos é muito mais profundo do que nos casos anteriores, o que não é apenas resultado deste tipo de linguagem, que por abranger, normalmente, um campo mais específico do que as classificações, particulariza com mais detalhe os conceitos, mas também consequência de uma maior aposta deste serviço nas potencialidades deste tipo de linguagem.

A comparação entre o nível de especificidade na representação do mesmo conceito nas três formas de *browsing* referidas confirma o que temos vindo a referir e traduz bem as diferenças assinaladas (Cf. Tabela 41).

A opção por uma organização alfabética dos descritores retirados, quer da DDC quer do CAB, se tem a vantagem da facilidade intuitiva, reduz claramente as potencialidades destes dois tipos de linguagens para o objectivo de disponibilizar o *browsing* das colecções.

Em termos de funcionalidades de navegação na estrutura e para o caso do Directório, o utilizador tem acesso inicialmente a uma representação da árvore que identifica todos os assuntos principais e dois ou três dos assuntos de primeiro nível. O critério para a selecção destes assuntos não é perceptível, a não ser a tentativa de eventualmente chamar a atenção dos utilizadores para os aspectos considerados mais relevantes. Assim, a este primeiro nível o utilizador fica com uma percepção reduzida da abrangência dos assuntos representados.

O desenvolvimento de cada um dos grandes temas é feito segundo um mesmo padrão: depois de seleccionar o tema pretendido, o utilizador tem acesso a uma lista de assuntos ordenados alfabeticamente, exceptuando os casos em que o assunto tem um carácter geral posicionando-se sempre em primeiro lugar. Quando os assuntos deste primeiro nível hierárquico não têm subdivisões é dada informação respeitante ao número de recursos atribuídos. Seleccionando um desses assuntos, o utilizador tem então acesso a uma lista de descrições de recursos, organizada alfabeticamente a partir do título. Quando os assuntos têm subdivisões, tal é indicado pelo sinal de mais (+) à frente de cada termo. Seleccionando esse assunto, o utilizador é conduzido a uma nova página onde se afixam os assuntos a esse nível. Ao longo de todo o processo, o utilizador vai tendo informação sobre o caminho percorrido na estrutura o que lhe permite, não só navegar na lógica da árvore, mas ter a noção permanente do contexto em que se insere o assunto em que num dado momento está posicionado. Esta informação indica todos os níveis anteriores, como se pode constatar pelo seguinte exemplo: *Home* → *Agriculture* → *Agencies and Organizations* → *International Agencies and Organizations*.

## Natural

Este serviço utiliza apenas a CDD para organizar os recursos, a partir de 5 grandes temas representados maioritariamente por 3 dígitos, o que traduz um nível de especificidade de representação geral. O objectivo é, aliás, de acordo com o próprio serviço proporcionar uma estrutura em árvore de nível generalista.<sup>76</sup> Cada um destes grandes temas é subdividido em dois ou três sub-níveis. A subdivisão a um quarto nível só se verifica em dois casos e está relacionada com a necessidade de individualizar a realidade britânica (Cf. Tabela 42).

Embora os dígitos da classificação não estejam visíveis para o utilizador, a estrutura de subdivisão dos assuntos na qual este navega respeita a organização hierárquica original. A prova mais cabal desta aplicação é o facto de serem identificados assuntos sem hiperligações no sentido de posicionar um assunto mais específico (já com hiperligações) no seu contexto. A análise exaustiva de cada uma das classes permite igualmente verificar a aplicação da lógica hierárquica da tabela aos casos seleccionados: todas as subdivisões de um assunto, em qualquer nível em que esteja posicionado, correspondem às subdivisões que a tabela implementa. Esta ligação à estrutura original

---

<sup>76</sup> Cf. Browse options. Disponível em <http://nature.ac.uk/browse/>

não impede a flexibilidade necessária de posicionar alguns assuntos ao mesmo nível mesmo quando eles são representados na tabela por um nível diferenciado. No entanto, nada disto é suficientemente discrepante para desvirtuar o sentido da utilização da CDD.

A tendência geral de representação dos assuntos desenvolve-se à volta das especificidades possíveis de representar pela utilização de dois dígitos após os três iniciais. O número quantitativo de níveis é relativamente reduzido, se comparado com outros serviços que temos vindo a analisar.

A distribuição dos recursos classificados em cada nível revela (Cf. Tabela 43) que é sobretudo no primeiro destes níveis que eles se encontram (66.2% dos recursos são aí classificados), sendo que mesmo no caso em que o assunto apresenta subdivisões várias existem sempre recursos classificados no nível mais geral, com diferenças quantitativas substanciais, como é o caso, por exemplo do assunto *Botany – Plants* (cf. Tabela 43)

Nem todas as temáticas representadas têm a mesma importância em termos do número de recursos nelas classificados. A distribuição do número de recursos pelos grandes temas (Cf. Tabela 44) revela o nítido privilégio de alguns assuntos (por exemplo *Zoology – Animals* e *Botany- Plants*) e a conflagrada representatividade de outros (por exemplo *Biogeography*, *Protistology* ou *Evolution*). Se estendermos esta análise para além do nível dos grandes temas verifica-se o mesmo tipo de distribuição, apresentando, por exemplo, os assuntos que contêm apenas um recurso classificado, uma percentagem importante de todos os assuntos representados (69.8%).

Embora o serviço disponibilize apenas 3.451 recursos, na estrutura de *browsing* encontram-se identificados 7.918, o que traduz uma alta taxa de atribuição na ordem dos 129%.

A navegação na estrutura é feita a partir de uma tabela que contém todos os assuntos principais, listados sem critério aparente, nem mesmo o alfabético. Para além da designação do assunto principal, o utilizador encontra ainda referência a 3 sub-assuntos, não sendo perceptível qual o critério que conduziu à sua inclusão. Após a selecção de um dos assuntos principais, o utilizador é conduzido a uma nova página onde se listam alfabeticamente todas as subdivisões desse assunto com todos os níveis e com a indicação do número de recursos classificados em cada nível.

Ao percorrer a estrutura de *browsing*, encontra-se sempre informação sobre o caminho percorrido, como por exemplo Top → *Evolution* → *Animals* → *Humans*. O facto de estes níveis serem hiperligações, permite ao utilizador, com facilidade, voltar a qualquer um dos níveis anteriores, mesmo à tabela inicial.

Este serviço não disponibiliza qualquer opção de pesquisa por palavra no contexto da estrutura de *browsing*.

## **Biores (Bio Research)**

Este serviço disponibiliza o acesso à colecção por duas vias: através da classificação desenvolvida pela *National Library of Medicine (NLM)* e pelo tesouro *Medical Subject Headings 2002 (MeSH2002)*<sup>77</sup>.

No primeiro caso, a aplicação da linguagem escolhida fica a um nível de representação generalista, aliás de acordo com o objectivo que o serviço pretende atingir com a utilização desta linguagem, que é o de proporcionar uma visão geral dos assuntos abordados. A análise das notações correspondentes a cada um dos descritores revelou (cf. Tabela 45), não sem alguma estranheza, que a grande maioria daquelas têm origem na LCC e não na NLM, utilizando inclusive a tabela QR da LCC, o que não é aconselhado pela NLM, como já foi assinalado. Assim, a opção pela NLM parece revelar-se pouco adequada.

A possibilidade de atribuir um recurso a mais do que um assunto é muito explorada, apresentando uma taxa de atribuição na ordem dos 46.8%.

A distribuição dos recursos pelos assuntos seleccionados é muito desigual, apresentando valores que oscilam entre 0.03% e 13.6%.

O nível de especificidade na representação dos conteúdos é muito maior na utilização do tesouro, que é apresentado alfabeticamente numa lista de termos a partir da qual o utilizador pode navegar. Este maior nível de detalhe resulta de uma opção clara do serviço. Os descritores são apresentados por ordem alfabética, sem qualquer subdivisão hierárquica, o que seria possível de implementar mesmo considerando as características desta linguagem. Assim, a não ser pelo nível de especificidade utilizado, não existe qualquer tipo de vantagem na utilização desta linguagem e do tesouro MeSH. O número de descritores utilizado é relativamente reduzido, o que se poderá dever à especificidade das temáticas objecto deste serviço.

## **VetGate**

Este serviço disponibiliza duas formas alternativas de acesso à colecção, através da utilização da Classificação da Biblioteca do Congresso e do tesouro CAB.<sup>78</sup>

No primeiro caso é disponibilizada uma lista alfabética de assuntos seleccionados da *LCC (Library of Congress Classification)*. Os assuntos são representados a um nível geral, sendo apenas 35 (em 91) claras subdivisões do assunto principal, apresentando por isso um nível de especificidade de representação mais alargado. No entanto, o facto de as notações LCC serem, como vimos, pouco expressivas relativamente à posição do assunto no contexto, a utilização de uma notação mais ou menos extensa, exceptuando nos casos assinalados, não representa uma maior especificidade de representação. Aliás, o próprio serviço esclarece que a LCC é aqui utilizada de forma generalista.

A utilização da possibilidade de atribuir um mesmo recurso a vários assuntos é pouco expressiva, apresentando a reduzida taxa de 3.8%. As razões desta baixa taxa de atribuição radicam no reduzido número de recursos que o serviço identifica (2016) e na opção por utilizar este nível a nível geral.

---

<sup>77</sup> Disponíveis em <http://bioresearch.ac.uk/browse>

<sup>78</sup> Disponíveis em <http://vetgate.ac.uk/browse>

Em termos da distribuição dos recursos pelos assuntos seleccionados verifica-se uma grande predominância de assuntos com relativamente poucos recursos classificados (Cf. Tabela 46). A única excepção a esta tendência é a dos assuntos onde se encontram classificados recursos no intervalo entre os 101-150. Estes assuntos apresentam as seguintes características, que justificam uma maior concentração de recursos neles classificados: são assuntos de uma importância central tendo em conta a abrangência temática do serviço – *Cattle diseases, Communicable diseases of animals, Horse diseases, Pet diseases, Veterinary medicine of special organs, regions and systems* – ou são assuntos de carácter muito generalista – *General veterinary medicine periodicals, Societies and serials, Veterinary study and teaching, research, research institutions and laboratories*.

## OMNI

Este serviço disponibiliza duas formas de acesso à colecção<sup>79</sup>, para além da pesquisa por palavra: duas listas alfabéticas de assuntos, uma com base na NLM que visa, de acordo com o próprio serviço, dar uma panorâmica geral dos assuntos cobertos, e outra com base nas MeSH, que representa os assuntos com maior especificidade.

A aplicação da NLM traduz-se na selecção de 84 descritores, apresentados sem a respectiva notação, e ordenados de forma alfabética. Nenhum destes assuntos apresenta qualquer subdivisão, o que, sem dúvida, contribui para a concentração de um número excessivo de recursos em muitos dos assuntos.

Como se verifica pelos dados recolhidos na Tabela 48 o número de assuntos com muitos recursos é significativo, o que certamente estará relacionado com a antiguidade relativa deste serviço em relação ao conjunto analisado e pela manutenção do mesmo nível generalista de representação apesar do crescimento da colecção. Os assuntos com um maior número de recursos atribuídos dizem respeito, como seria de esperar, a temáticas centrais na área da saúde, basicamente tudo o que diz respeito às diversas especialidades médicas, saúde pública, hospitais e formação.

A alta taxa de atribuição (52%) vem reforçar a ocorrência deste fenómeno, com as evidentes dificuldades de navegação que cria aos utilizadores.

O acesso à colecção através das MeSH apresenta-se igualmente sob a forma de uma lista alfabética de descritores, subdividida pelas letras do alfabeto.

No que concerne às funcionalidades de navegação, no caso da NLM, o utilizador selecciona o assunto pretendido consultando a lista alfabética de assuntos, após o que é conduzido a uma página que apresenta a totalidade dos recursos classificados ordenados pela ordem alfabética dos títulos. Apenas a hiperligação associada ao nó “up” lhe permite voltar à lista geral de assuntos.

Utilizando os descritores retirados das MeSH, o utilizador pode seleccionar o assunto pretendido através da consulta da lista de assuntos. Após a selecção do assunto, acede a uma página com a lista total de recursos e com a indicação, quer da posição relativa do

---

<sup>79</sup> Disponíveis em <http://omni.ac.uk/browse>

termo na relação hierárquica (são disponibilizadas hiperligações para todos os termos gerais e específicos no qual o termo seleccionado se integra), quer da sua relação com outros termos.

Através da hiperligação associada ao termo “*up*” é possível voltar à lista inicial directamente para a página da letra do alfabeto em que o termo de partida se insere. Nos casos em que são usados qualificadores, o conjunto de termos que resulta da aplicação não está disponível a partir da lista principal, mas apenas quando se acede ao respectivo assunto principal e nestes casos a lista de recursos aparece subdividida de acordo com o conjunto de termos que resulta da utilização dos referidos qualificadores. Exemplifiquemos: os vários descritores associados ao conceito de “*skin*” aparecem alfabetados na letra “S”, mas os termos “*Skin/pathology*” e “*Skin/radiation effects*”, que resultam da aplicação dos qualificadores “*pathology*” e “*radiation effects*” só estão disponíveis após aceder à lista de recursos para o termo “*Skin*”. Não existe qualquer razão, nem ligada à lógica da linguagem nem à forma de organização do acesso, que impeça a presença destes termos construídos na respectiva posição alfabética da lista principal. Tal poderia justificar-se se não existissem recursos classificados com o termo mais geral (“*Skin*”) e só existissem recursos para os termos com a adição dos qualificadores, mas tal não se verifica.

#### 4.3.2.5. EEVL

O recente alargamento temático do serviço às áreas da Matemática e da Computação e o facto do desenvolvimento destes assuntos estar cometido a entidades diferentes, embora parceiras, justificará talvez a distinção das três áreas para o acesso por *browsing*. A relativa autonomia das temáticas, se bem que com bastantes pontos de contacto, poderá eventualmente justificar estruturas de *browsing* não unificadas considerando que o utilizador, quando aborda este serviço tem já definida, pelo menos, a grande área temática em que pretende pesquisar. Qualquer que seja a razão invocada, o facto indiscutível é o de que, graças a esta implementação, o utilizador jamais poderá navegar numa estrutura conjunta.

Cada uma das áreas temáticas deste serviço é organizada para efeitos de *browsing* dos recursos através de três linguagens diferentes: a *Mathematics Subject Classification 2000* (MSC 2000) para a área da matemática, a *ACM computing classification system* (1998) para a área da computação e os Tesouros e códigos de classificação *Ei* (*Engineering Information*).

No caso da *ACM computing classification*, a linguagem é utilizada exactamente com a estrutura original, quer ao nível das grandes divisões e de todas as subdivisões (Cf. Tabela 49). Em alguns casos suprimem-se subdivisões de 2º ou 3º nível. Não se verifica qualquer caso de expansão da classificação original. A utilização contempla mesmo os descritores originais, embora excluindo a notação. Foi retirada em todas as classes a subdivisão “XXm = Miscellaneous”

No caso da MSC 2000, as subdivisões principais de EEVL não correspondem às classes principais da classificação. São definidas pelo serviço e representam a criação de um nível superior, mais generalista, do que o nível utilizado na classificação. Por exemplo, os assuntos *Combinatronics*, *Order*, *lattices*, *ordered algebraic structures*, *Number*

*theory, Field theory and polynomials, Commutative rings and álgebras, Algebraic geometry, Linear and multilinear álgebra; matrix theory, Associative rings and álgebras, Nonassociative rings and álgebras, Category theory; homological álgebra, K-theory, Group theory and generalisations, topological groups, Lie groups e General algebric systems* todas classes principais na MSC aparecem aqui num segundo nível dependente de um nível mais geral designado por *Algebra*.

O nível de representação é generalista, fica-se pelas classes principais da MSC. São raras as situações em que assuntos representados a um nível inferior de especificidade na MSC são colocados a um nível superior na estrutura de *browsing*. Exemplo: Mathematics departments and institutions que na MSC tem a notação 01A74, fica posicionado no segundo nível da estrutura. Também este caso constituiu o único exemplo de uma extensão realizada por este serviço à classificação básica, especificando os departamentos e instituições matemáticos através de uma subdivisão geográfica por países.

Os assuntos são subdivididos em 4 níveis máximos, maioritariamente em três, sendo o nível de especificidade máximo na representação dos assuntos, mas, como já se referiu, este nível representa os assuntos a um nível ainda de grande generalidade. São frequentes, por exemplo, mesmo no terceiro nível, as formulações que associam mais do que um conceito ou facto no mesmo termo classificador: *Electrical power, sources and motors; Machinery & Machine tools; Project and people management; Testing, certification and licensing*. No entanto, a diversidade dos assuntos representados é significativa, encontrando-se o assunto *Engineering* em primeiro lugar logo seguido do tema *Computing*, que aliás constituem os temas claramente privilegiados se considerarmos que o tema *Mathematics* contém apenas 24.7% de todos os assuntos representados (Cf. Tabela 50).

O posicionamento em primeiro lugar do assunto *Engineering*, bem como um maior nível de especificidade no desenvolvimento ficam a dever-se à participação de um *gateway* especializado nas ciências aeroespaciais e na defesa designado por AERADE<sup>80</sup>. Praticamente toda a estrutura de *browsing* utilizada por este último passou para EEVL: os assuntos não utilizados dizem respeito a aspectos que caem fora do âmbito temático de EEVL, tais como, por exemplo, história militar ou relações internacionais. No entanto, quando comparados sistematicamente os dois desenvolvimentos, em AERADE e EEVL, verificam-se excepções difíceis de justificar. Vejamos alguns exemplos: no assunto *Military vehicle technology* dos 6 sub-assuntos em AERADE apenas foram seleccionados 3; no assunto *Weapon Technology* também não é perceptível o critério que conduziu à não inclusão em EEVL de *Rocket systems, Land Mines e Ground-based artillery*. A razão destas relativas incongruências pode estar relacionada com uma eventual *décalage* temporal entre a actualização de AERADE e de EEVL.

Embora só descreva 9.700 recursos *Web*, na estrutura de *browsing* é possível encontrar consideravelmente mais recursos na ordem dos 20.793. Trata-se da ocorrência do fenómeno atribuição, que neste caso atinge a elevada taxa de 122.7%.

Apesar de ser no terceiro nível hierárquico que existe um maior número de assuntos, é sobretudo ao segundo nível que os recursos são classificados, com uma diferença

---

<sup>80</sup> AERADE: portal to quality aerospace and defence resources on the Internet. Disponível em: <http://aerade.cranfield.ac.uk/>

significativa em relação ao nível seguinte: enquanto que os recursos classificados no segundo nível representam 50.6% de todos os recursos presentes na estrutura, os classificados ao terceiro nível representam apenas 33.5% (Cf. Tabela 51). Considerando a alta taxa de atribuição e o facto de os recursos no segundo nível terem necessariamente um carácter mais generalista do que os que estão presentes no terceiro, esta significativa diferença encontrará, pelo menos parte, do seu significado.

O nível de especificidade da classificação dos recursos acompanha o nível de especificidade da representação do assunto na linguagem, quer dizer, não existem recursos classificados a um nível mais geral quando esse nível se subdivide noutros mais específicos. As únicas excepções a esta regra são os assuntos: *Aeronautics*, *Aerospace structures and materials*, *weapon technology* e *missile technology* do segundo nível; e *Aeronautics (General)*, *Aircraft Manufacturing, production and maintenance*, *Electronic Warfare* no terceiro nível. Todas as excepções se verificam no tema *Aerospace and Defence engineering*, área desenvolvida, como já vimos, por AERADE. As razões deste fenómeno são várias: em primeiro lugar, todos os recursos neste tema que não vêm directamente de AERADE são posicionados no tema mais geral; em segundo lugar existem erros de atribuição do recurso ao nível exacto em que se deve situar, por exemplo, o recurso com o título “ALLVAC” está classificado em EEVL no nível 2 (*Aerospace Structures and Materials*) enquanto que em AERADE, o mesmo recurso está classificado no nível 3 (*Metals*); finalmente, em alguns casos descrições vindas de AERADE são classificadas a um nível mais geral mesmo quando esse nível tem subdivisões, não se recorrendo à estratégia de criar uma subcategoria do tipo “General” como muitas vezes é feito em EEVL (Exemplo: o tema *Electronic warfare* tem dois recursos classificados, quando tem a subdivisão *Electronic countermeasures*).

O cálculo do número de recursos presente na estrutura permite igualmente perceber que existe um grande desequilíbrio no que diz respeito à representatividade das temáticas globais. Assim, enquanto que 72.4% dos recursos dizem respeito ao tema *Engineering*, os outros dois temas quedam-se por valores à roda dos 14% (13.7% para *Mathematics* e 13.8% para *Computing*).

A grande maioria dos assuntos apresenta um reduzido número de recursos classificados, o que traduz uma das principais desvantagens destes serviços, muitas vezes referida pelos utilizadores e que é a de identificarem relativamente poucos recursos se comparado com outros. Não obstante, pode sempre argumentar-se que se tratam de registos submetidos a um apertado crivo de qualidade e esta faceta compensará largamente a reduzida expressão quantitativa. De facto, a grande maioria dos assuntos tem entre 1 e 10 recursos identificados (Cf. Tabela 52).

A concentração dos valores à volta deste intervalo e do seguinte, representando, 60% dos recursos aponta decisivamente para uma reduzida expressão do número de recursos em cada assunto.

Se partirmos do pressuposto de que um assunto que contenha apenas até 5 recursos classificados é um assunto escassamente representado, com a óbvia probabilidade de frustrar o utilizador, e se contabilizarmos o número de assuntos nestas condições verifica-se que os valores são significativos, representando 27.8% de todos os assuntos (Cf.

### Tabela 53).

Por outro lado, o valor relativamente significativo de assuntos com mais de 100 recursos classificados (12.3%) obriga o utilizador a percorrer um número significativo de páginas para seleccionar os recursos que melhor se adequam às suas necessidades. No assunto com o maior número de recursos classificados – *Engineering General* – com 640 recursos o utilizador tem que percorrer 32 páginas, e mesmo num assunto com 104 recursos classificados, já é obrigado a percorrer 6 páginas de descrições. A dificuldade em seleccionar os recursos apropriados através desta forma de visualização é reveladora de uma das principais limitações deste modelo *browsing*, ou seja, quanto mais recursos se classificam em cada assunto mais difícil é para o utilizador seleccionar o ou os recursos adequados. A resolução deste problema pode passar eventualmente por dois tipos de soluções, que não se excluem necessariamente. A primeira seria a de desenvolver a especificidade de representação acrescentando sucessivamente níveis mais específicos de representação dos conceitos; a grande desvantagem é a de que o utilizador se veria obrigado a percorrer muitos níveis da árvore até chegar àquele que verdadeiramente lhe interessa. A segunda abordagem possível é a de disponibilizar a pesquisa por palavra em cada um dos níveis de assunto. Se bem que esta solução possa parecer contraditória com o tipo de comportamento de pesquisa a que estas forma de organização visa responder, sobretudo se tivermos em conta que uma das principais dificuldades do modelo de pesquisa por palavra é da formulação da pergunta, o utilizador pode com facilidade chegar à formulação de uma pergunta através de uma rápida análise de uma amostra dos recursos classificados. É esta última abordagem que este último serviço que estamos a estudar implementa. Em cada um dos níveis da estrutura de *browsing* o utilizador tem disponível a pesquisa por palavra dentro dessa categoria ou em todo o serviço, com as mesmas funcionalidades de restrição ou expansão já descritas.

Para além deste aspecto, o utilizador tem constante informação sobre o caminho que percorreu na estrutura, podendo recuar para qualquer um dos níveis anteriores até ao topo, ou seja, ao tema geral. Assim, se o utilizador entrar no serviço por um ponto mais específico dessa estrutura é-lhe fornecida informação sobre o contexto em que o assunto se insere no âmbito da estrutura. No entanto, apesar de existirem, como vimos, ligações de tipo associativo entre os termos nestas linguagens, nenhuma dessas ligações é representada na estrutura de *browsing*. A aposta na relação hierárquica como base de construção da estrutura não impediria a sinalização de outro tipo de relações, podendo, ao contrário acrescentar uma evidente mais valia para o utilizador. As remissivas de exclusão seriam particularmente adequadas neste contexto por providenciarem formas alternativas de designação de um assunto.

A informação sobre o número de recursos ocorre sempre nas categorias onde existem recursos classificados. Para saber o número de recursos total de uma categoria que seja subdividida, o utilizador terá de somar os valores apresentados em cada uma das subcategorias dessa categoria.

Na globalidade verifica-se que estes serviços utilizam dois grandes tipos de linguagens documentais – classificações e tesouros – para proporcionar um acesso à colecção

diferenciado correspondendo a comportamentos de pesquisa também eles diversos. Assim, as classificações são utilizadas para permitir uma visão abrangente mas generalista dos conteúdos, o que é particularmente adaptado ao comportamento de pesquisa que caracterizamos como semi-definido. As classificações são particularmente adequadas a este desiderato, na medida em que permitem representar os assuntos a diferentes níveis de especificidade, podendo portanto optar-se por um nível de maior generalidade, o que permite reunir um conjunto relativamente significativo de recursos para aquele tema. O recurso aos tesouros, que aponta para uma representação de grande especificidade, permite ir ao encontro de utilizadores que identificam já claramente o termo ou termos que podem corresponder à satisfação das suas necessidades, ou seja, revelam um comportamento já estruturado.

Não obstante, e sobretudo em relação às classificações, o nível de generalidade é diverso de serviço para serviço: a montante temos situações em que o nível de generalidade é muito alto ficando-se por representações ao nível do topo temático; a jusante, temos desenvolvimentos que vão ao ponto de especificar com detalhe as divisões geográficas (por exemplo SOSIG). Relativamente à CDU, o nível de especificidade depende igualmente da opção de utilização: a utilização de notações simples ou notações pré-coordenadas como notações simples, mas sobretudo no primeiro caso, determina um nível de especificidade restrito na medida em que cada vez mais esta linguagem apresenta uma estrutura mais facetada (SALVIC:2003). Assim, à medida em que se aumentar a especificidade mais necessidade existirá de utilizar combinações de notações. Tal é o caso do desenvolvimento de alguns assuntos no serviço SOSIG. Esta forma de utilização da CDU aproxima-se das características mais enumerativas quer da CDD, quer da LCC ou da NLM. Aliás, a utilização da CDU desta forma mais enumerativa ultrapassa as dificuldades da utilização de notações sintéticas e pré-coordenadas, se bem que se percam as possibilidades de uma representação mais rica em termos de especificidade.

Quer num quer noutro caso, a grande maioria dos serviços utiliza linguagens documentais tradicionais, quer dizer, com grandes tradições no ambiente dos recursos em papel. A opção por estas linguagens pode justificar-se não só pela economia dos processos evitando assim os custos de produção de uma nova linguagem, mas também por outras razões que se prendem com a adequação destas linguagens aos objectivos que se pretendem atingir e com o conhecimento técnico instalado sobre a sua utilização. De facto, como tivemos oportunidade de realçar, quer do ponto de vista teórico quer do ponto de vista aplicacional as classificações apresentam diversas virtualidades particularmente ajustadas ao ambiente em causa e ao objectivo de proporcionar uma visão abrangente do conteúdo dos serviços.

Do ponto de vista da implementação das classificações verifica-se, em primeiro lugar, a presença de dois modelos básicos: um que podemos designar por hierárquico e outro por alfabético. No primeiro caso, os descritores seleccionados são apresentados de acordo com a lógica hierárquica do geral para o particular, mantendo-se mesmo em alguns casos assuntos sem qualquer recurso classificado de forma a contextualizar, na lógica da linguagem, os assuntos identificados. No segundo caso, os descritores são apresentados apenas de forma alfabética. A opção por estas duas diferentes formas de implementação pode estar directamente relacionada com a especificidade temática do serviço. Serviços com temáticas mais abrangentes optam por uma estrutura com vários níveis de especificidade, permitindo assim representar uma realidade que ela própria se

desmultiplica em vários níveis de abordagem. Serviços com temáticas mais específicas optam por seleccionar os assuntos de acordo com essa maior especificidade, não atribuindo qualquer importância à sua contextualização na estrutura hierárquica. A opção por este modelo alfabético revela-se, no entanto, desadequada se tivermos em linha de conta que os assuntos aparecem descontextualizados, perdendo assim a possibilidade de dar a relação hierárquica, transmitindo-se a percepção de que se trata de um conjunto de assuntos que se encontram ao mesmo nível de generalidade; por outro lado, o crescimento do número de assuntos identificados obrigará necessariamente a percorrer uma lista alfabética cada vez mais extensa sem qualquer nível intermédio de selecção.

Duas outras variáveis podem influenciar a opção por este modelo: o nível de generalidade e o tipo de classificação escolhida. Se o nível de generalidade escolhido for muito alto, quer dizer ao nível das grandes classes, não faz sentido criar uma estrutura de subdivisões hierárquicas, pelo que a ordenação alfabética dos grandes temas aparece como solução mais intuitiva. A utilização de classificações mais enumerativas pode também contribuir para a opção por este modelo, se bem que estamos sempre em presença de uma estrutura de assuntos de acordo com o princípio organizativo do geral para o particular; o facto de a notação poder não traduzir completamente as subdivisões só em pequena monta justificaria uma opção deste tipo. Finalmente, se compararmos esta solução com a implementada para a organização utilizando tesouros, verifica-se que, exceptuando o maior nível de especificidade do segundo caso, nada distingue um e outro tipo de organização, pelo que ainda menos se justificaria a opção por dois tipos de linguagens diferentes.

Mesmo que o nível de generalidade seja relativamente alto, e nunca o é ao nível das grandes classes apenas, a opção por organizar hierarquicamente os assuntos apresenta vantagens assinaláveis. Vejamos então um exemplo de organização hierárquica para um serviço que utiliza actualmente o modelo alfabético: o caso OMNI. Neste serviço, o número de assuntos retirados da NLM é já significativo (na ordem dos 90) e o nível de generalidade é alto na grande maioria dos casos. A aplicação directa dos níveis da tabela revelou um excesso de subdivisões desnecessárias, quer por não terem recursos associados, quer por não existirem outras subdivisões que sejam consideradas pelo serviço. Assim, optou-se por elaborar uma proposta adaptada mantendo o princípio hierárquico, que a seguir se apresenta:

### **Experiência de organização hierárquica para *browsing* de OMNI (Adaptada)** (alfabetada a partir dos códigos das classes)<sup>81</sup>

#### **(QS) *Human Anatomy***

(QS4) *Anatomy (General)*

(QS504) *Histology (General)*

(QS604) *Embryology (General)*

#### **(QV) *Pharmacology***

(QV4) *Pharmacology (General)*

(QV600) *Toxicology (General)*

(QV700) *Pharmacy*

---

<sup>81</sup> Foram apenas utilizados descritores retirados da NLM, excluíram-se portanto todos os que têm origem na LCC

**(QW) Microbiology. Immunology**

- (QW4) Microbiology (General)*
- (QW504) Immunology (General)*

**(QZ) Pathology**

- (QZ4) Pathology (general)*
- (QZ200) Neoplasms*

**(W) Health professions**

- (W1) Medical journals and other general publications*
- (W3) Congresses, conferences*
- (W20.5) Medical research in general, sources of funding*
- (W21) Professional information*
- (W26) Equipment and supplies, catalogues (General)*
  - (W26.5) Informatics, computing*
- (W30) Law, jurisprudence, medical ethics*
- (W74) Health economics*
- (W83) Telemedicine*
- (W84) Health services*
- (W89) Family practice*
- (W601) Forensic medicine*

**(WB) Practice of medicine**

- (WB102) Practice of medicine, general diagnosis and therapeutics*
- (WB300) Therapeutics (General), including alternative medicine*
- (WB320) Rehabilitation, disability*
- (WB400) Dietetics*
- (WB460) Physical therapy*
- (WB555) Occupational therapy*

**(WC) Communicable diseases**

- (WC100) Communicable diseases (General)*
- (WC503) Acquired immunodeficiency syndrome*

**(WD) Disorders of systemic, metabolic or environmental origin, etc.**

- (WD100) Nutrition, nutrition disorders (General)*
- (WD200) Metabolic diseases (General)*
- (WD300) Immunologic and collagen diseases. Hypersensitivity.*
- (WD400) Animal poisons (General)*
- (WD500) Plant poisons (General)*
- (WD700) Aviation and Space medicine (General)*

**(WE) Musculoskeletal system**

- (WE100) Musculoskeletal system (General)*
- (WE168) Orthopaedics*
- (WE890) Podiatry*

**(WL) Nervous system**

- (WL100) Nervous system (General)*

*(WL340.2) Speech therapy, communicative disorders*

***(WN) Radiology. Diagnostic imaging***

*(WN100) Radiology (General)*

*(WN180) Radiography, diagnostic imaging (general)*

***(WO) Surgery***

*(WO100) Surgery (General)*

*(WO200) Anaesthetics (General)*

***(WQ) Obstetrics***

*(WQ100) Obstetrics (General)*

*(WQ160) Midwifery*

***(WY) Nursing***

*(WY19) Nursing schools*

*(WY100) Nursing (General)*

*(WY106) Health visiting, community health nursing*

Esta experiência demonstra, que o utilizador ganha informação de enquadramento dos assuntos, que não pode ser dada por uma organização alfabética. Por outro lado, em termos de crescimento futuro uma tal organização permite a adição de novos assuntos sem obrigar o utilizador a percorrer uma extensa lista alfabética. Por último, se tivermos em linha de conta o objectivo com que são usadas estas classificações, então a presença das categorias mais gerais permite ao utilizador menos experimentado uma percepção ainda mais global das temáticas cobertas pelo serviço.

A opção por um maior ou menor nível de especificidade produz, como é evidente, resultados diversos a vários níveis, demonstrando as potencialidades e limitações da aplicação destes instrumentos organizativos.

Ao nível do número de recursos classificados em cada categoria, a opção por um nível de maior generalidade traduz-se tendencialmente na concentração de um maior número de recursos em cada um dos assuntos. Tal apresenta a clara desvantagem de reduzida discriminação entre os recursos, de obrigar o utilizador a percorrer extensas listas de descrições até identificar os itens de interesse. Estas desvantagens aumentam se se associarem, sem contextualização hierárquica, assuntos de níveis de especificidade maior produzindo situações de grande disparidade, que certamente não contribuem para o almejado objectivo de proporcionar uma visão clara do serviço.

Optando pelo desenvolvimento mais específico da árvore de assuntos, verifica-se que um aumento do nível de especificidade conduz tendencialmente a um aumento dos assuntos com um número reduzido de recursos classificados: a correlação entre o número de níveis de cada serviço e o número de assuntos com pouco recursos classificados (até 5) apresenta o valor significativo de +0.91. O factor tempo influencia o desenvolvimento desta tendência: nos serviços mais antigos (vide, por exemplo, SOSIG) o aumento do número de recursos produz um correspondente aumento do número de níveis de especificidade, bem como a atribuição de um maior número de recursos em cada nível. Verifica-se assim que uma estrutura hierárquica mais rica só

ultrapassará a desvantagem de um reduzido número de recursos em cada assunto se a colecção for quantitativamente significativa. No entanto, o desenvolvimento da hierárquica apresenta sempre este perigo, ou seja, o de existirem um significativo número de assuntos com poucos recursos. Uma das formas de ultrapassar esta desvantagem pode ser a de permitir ao utilizador aceder aos recursos do nível específico e/ou a todos os recursos classificados a um nível mais geral e a todos os seus sub-níveis. Uma outra desvantagem que podemos antever em estruturas de grande especificidade é a de obrigar o utilizador a navegar sucessivamente num extenso número de níveis até encontrar o tópico de interesse. A vantagem de aumento da especificidade é, por um lado, a de anular a desvantagem já identificada da excessiva generalidade e, por outro, a de se adaptar a necessidades mais diversificadas dos utilizadores.

A grande maioria dos serviços desenvolveu uma grande flexibilidade na aplicação da linguagem em termos da representação da necessária especificidade dos assuntos, destacando temas que se encontravam eventualmente “escondidos” num nível mais específico na estrutura original. Isto foi conseguido, quer pela anulação dos níveis mais gerais, o que produz a desvantagem da descontextualização, quer pela utilização da mesma notação (mais geral) para diferentes assuntos, aproveitando assim as capacidades de síntese da classificação. Muitas vezes esta opção pela especificidade não corresponde, como já se verificou, a uma colecção necessariamente desenvolvida.

A possibilidade de atribuir um mesmo recurso a vários assuntos é bem explorada por quase todos os serviços. Exceptuando OMNI, BioResearch, PSIGATE, VetGate e Agrifor, todos os outros apresentam taxas de atribuição superiores a 100%.

A aplicação das classificações parece orientar-se por um modelo com duas alternativas possíveis: manter um nível de alta generalidade com o inconveniente de perda de discriminação e de resultar num excessivo número de recursos em cada tema, o que torna difícil a navegação do utilizador; desenvolver a especificidade com o risco de apresentar um número elevado de assuntos com poucos recursos. A solução destas questões parece passar por uma gestão da estrutura hierárquica que avalie periodicamente a necessidade de criar novos níveis de especificidade quando o número de recursos o justificar ou eventualmente anulá-los quando esse número diminuir. Até lá e para evitar desnecessárias e sucessivas reclassificações, deve optar-se pela atribuição da notação mais específica que, em termos de geração da árvore de assuntos deve posicionar o recurso nessa notação ou na mais geral se a anterior não existir. Uma outra forma de ultrapassar estes problemas é proporcionar o acesso alternativo de pesquisa por palavra-chave para além da navegação na árvores, como, aliás, já foi implementado por alguns dos serviços analisados ( vide, por exemplo, SOSIG e EEVL).

Ao utilizar um modelo alfabético simples na implementação de Tesouros, perdem-se todas as possibilidades relacionais, quer associativas quer hierárquicas. Todos os serviços que optaram por esta solução só permitem a percepção das relações entre os termos depois de seleccionar os referidos termos, o que apresenta, como é evidente, a vantagem de permitir ao utilizador navegar nas relações temáticas, mas não compensa totalmente o inconveniente de ter de percorrer uma lista alfabética de termos sem qualquer contextualização. A utilização deste modelo de implementação dos Tesouros só poderá ser realizada com alguma vantagem por indivíduos que se apresentem perante esta opção com um comportamento de pesquisa bem definido.

### 4.3.3. Descrições em texto livre

A Descrição constitui, como vimos, um dos elementos obrigatórios de todas as estruturas de metadados nos serviços em análise. A importância que lhe é conferida é grande, sendo um dos campos de informação sempre mostrados mesmo nas visualizações mais minimalistas. A importância conferida a este elemento decorre das funções que pode desempenhar em termos de selecção dos recursos pelos utilizadores.

Em todos os serviços (Cf. Tabela 54), a semântica deste elemento é descrita com algum cuidado, até pela sua natureza distinta de todos os outros elementos, ou seja, é o único cujo conteúdo não pode ter uma grande pré-determinação. As directrizes para a produção de resumos apresentam alguma semelhança entre os vários serviços, observando-se mesmo uma igual formulação para PSIGATE e BIOME (em todos os seus componentes, ou seja, AGRIFOR, BIORES, BIOME, NATURAL, VETGATE e OMNI).

Em termos de extensão do resumo, só o serviço SOSIG não indica um número mínimo ou máximo de palavras. HUMBUL aponta para uma média de mais ou menos 100 palavras com um limite máximo de 300. PSIGATE e BIOME apresentam um valor entre as 100 e as 150 palavras.

Poucas indicações são fornecidas relativamente à estrutura do resumo. HUMBUL aponta para a produção de textos pouco estruturados, ou seja, sem parágrafos e evitando longas enumerações em forma de lista. O único serviço a impor uma clara e inultrapassável restrição formal em termos de estrutura é SOSIG, visto que o software aplicativo utilizado para a inserção da informação permite apenas um único parágrafo. Todos os outros serviços não estabelecem directrizes quanto à estrutura do resumo.

Estabelecer uma regra quanto à repetição da informação da fonte é apenas preocupação clara de HUMBUL, afirmando especificamente “Não copie e cole a descrição directamente do sítio *Web*” (Humbul: cataloguing guidelines).

Como seria de esperar as directrizes mais desenvolvidas dizem respeito aos aspectos temáticos do conteúdo do resumo, que podem ser agrupadas em três grandes categorias: cobertura temática do recurso, informação sobre responsabilidade e tipologia do recurso.

No que diz respeito a aspectos mais específicos e eventuais limitações à utilização da informação vários são os aspectos referidos. Todos consideram a necessidade de referenciar a existência da mesma informação em outras línguas para além da original, sendo que em PSIGATE e BIOME este aspecto é restringido a outras línguas que não o inglês. A informação sobre limites geográficos e temporais do recurso é referida explicitamente por HUMBUL e SOSIG mas não por PSIGATE / BIOME. Isto poderá explicar-se pelo facto da variável tempo e espaço ter uma maior importância nos primeiros e uma reduzida relevância nos segundos considerando as temáticas que cobrem. Quanto a formatos específicos em que a informação se encontra só BIOME / PSIGATE referem especificamente este aspecto, embora HUMBUL ao estabelecer a necessidade de dar informação sobre a forma de apresentação do recurso possa também incluir a característica específica do formato. Só HUMBUL não inclui qualquer

referência explícita a condicionantes tecnológicas e económicas no acesso à informação. O estabelecimento de ligações com outros recursos é explicitamente referido por BIOME / PSIGATE, mas HUMBUL remete esta informação para o elemento Relação.

Nenhuma directriz significativa é dada quanto ao estilo (informativo, indicativo ou crítico seriam, pelo menos, as grandes categorias que poderiam ser invocadas), aspectos gramaticais, escrita (por exemplo, quanto à utilização de abreviaturas e numerais).

Em termos de directrizes mais generalistas, destacam-se as recomendações relativas à necessidade de síntese, chegando mesmo HUMBUL a aconselhar a concentração do essencial da descrição nas primeiras informações fornecidas sobre o registo. O apelo à objectividade, se bem que nem sempre explícito em todos os serviços, transparece em todos eles, com mais incidência para os *Gateways* da área das ciências sociais.

As directrizes fornecidas pelos serviços para a construção da descrição, se bem que sendo mais específicas do que qualquer um dos outros elementos, são na generalidade pouco esclarecedoras quanto a alguns aspectos essenciais na construção de resumos. Em primeiro lugar, definição clara da tipologia de resumo a implementar é deficiente, o que seria importante, já que, como tivemos oportunidade de ver, a definição de uma tipologia é decisiva para a identificação das várias categorias temáticas que o resumo deve conter, dado que cada tipologia aponta para o realce de determinados aspectos do conteúdo da informação. A estrutura do resumo é parâmetro praticamente esquecido, a não ser algumas referências gerais. Directrizes quanto à grafia aconselhada para determinadas palavras (acrónimos e numerais), bem como no que respeita à utilização da voz passiva são completamente ignoradas.

Apesar de serem considerados alguns aspectos da caracterização dos recursos informativos que estão intimamente ligados ao ambiente onde são disponibilizados, verifica-se que determinados aspectos bastante relevantes são completamente ignorados, por exemplo, os que dizem respeito à estrutura dos sítios e às formas de navegação. A ausência de directrizes relativamente a este aspecto releva de outros dois problemas mais gerais e que são, por um lado, a ausência de directrizes que considerem as diferentes características das diversas tipologias de recursos (sítios *Web*, *mailing lists*, etc.) e por outro a não consideração da variável granularidade da descrição.

Apenas as directrizes utilizadas por BIOME e PSIGATE têm em conta a tipologia de recursos, e mesmo assim só destacam o caso das *mailing lists*, estabelecendo que nesse caso os elementos de informação a considerar são: público-alvo, país ou área geográfica de origem e instruções sobre como aderir à lista.

Existem, por outro lado, nestas directrizes quanto ao conteúdo do elemento descrição alguns aspectos que parecem contraditórios, na medida em que implicam a repetição de informação, de forma não normalizada, que está já presente em outros elementos da estrutura de metadados. O exemplo mais flagrante é o da responsabilidade sobre o conteúdo: todas as estruturas em análise incluem o elemento responsabilidade, algumas até com vários qualificadores possíveis e estabelecem regras que asseguram a normalização da sintaxe do conteúdo desse elemento, pelo que a repetição do mesmo tipo de informação no resumo não se justifica, nem mesmo por motivos de recuperação da informação. O mesmo se passa para a informação respeitante à tipologia do recurso.

Após esta análise das directrizes do serviço quanto ao conteúdo do elemento em análise, todas as descrições da amostra seleccionada foram submetidas a testes de legibilidade já referidos, a saber: número de palavras, número de caracteres, número de parágrafos, número de frases, média do número de frases por parágrafo, média do número de palavras por frase, média do número de caracteres por palavra, percentagem das frases na voz passiva, Flesh Reading Ease e Flesh Kincaid Grade Level (cf. Ponto 4.1)

O primeiro aspecto a ter em conta é o da extensão do resumo, medida em número de palavras. Através da Tabela 55, que regista a média do número de palavras da amostra de cada serviço, é possível, em primeiro lugar, concluir que se tratam de resumos de extensão média (média = 102 palavras).

No entanto, os valores apresentam uma importante percentagem de dispersão em relação à média (coeficiente de variação = 38%), o que dá bem a ideia da heterogeneidade dos textos produzidos quanto à sua extensão. A hipótese de que os resumos na área das ciências sociais seriam tendencialmente mais longos, quanto mais não fosse pela complexidade de definição de muitos conceitos, não obtém uma absoluta confirmação no contexto de estudo, apesar de que HUMBUL apresenta valores muito superiores à média. Se considerarmos que a extensão dos resumos em SOSIG pode ser claramente limitada pela estrutura obrigatória de um único parágrafo e ainda que a reduzida distância que o separa de EEVL, poderíamos eventualmente considerar que os valores de SOSIG poderiam atingir uma média superior à de EEVL. E assim, de facto, teríamos os resumos que abordam temáticas das ciências sociais posicionados em primeiro lugar. A aplicação das mesmas directrizes (casos de BIOME e PSIGATE), por outro lado, produz resultados semelhantes, apesar das diferenças nas temáticas abordadas, se bem que esta diferença seja menor da que opõem, por exemplo, os valores extremos da tabela.

Se confrontada com as directrizes estabelecidas quanto a este parâmetro, a prática revela um afastamento significativo dos serviços PSIGATE e BIOME em sentido inferior ao menor valor aconselhado. As descrições em HUMBUL afastam-se ligeiramente da média aconselhada, mas estão longe de atingir o máximo. A significativa discrepância da prática em BIOME e PSIGATE significará uma menor atenção conferida a este elemento ou traduzirá o carácter mais sintético da linguagem científica nas ciências ditas exactas?

Considerando que palavras mais pequenas são indicativas de maior legibilidade, o parâmetro média do número de caracteres por palavra constitui um bom indicador para avaliar a complexidade das palavras (WHEATLEY:1997). Os dados constantes da Tabela 56 permitem concluir, em primeiro lugar, por uma grande homogeneidade entre todos os serviços.

Ao contrário da situação anterior, a média é, neste caso, mais significativa com o valor de 5.5, dada a reduzidíssima variação (coeficiente de variação = 1.8%).

Os valores um pouco mais elevados dos serviços PSIGATE e EEVL estarão certamente relacionados com a complexidade específica de um grande número de palavras nas áreas científicas respectivas.

Na generalidade, podemos afirmar que estamos perante valores que denotam uma complexidade média das palavras utilizadas, embora se deva levar em linha de conta

que os resultados desta medida devem ser encarados com reserva na medida em que a familiaridade dos utilizadores com certa terminologia pode ultrapassar as dificuldades causadas por uma maior extensão e vice-versa, alguns assuntos requerem a utilização de palavras mais complexas que outras e, por último, é preciso ter em conta que este tipo de teste não distingue adequadamente entre palavras e outros conjuntos de caracteres como acrónimos por exemplo (WHEATLEY:1997).

O cálculo do número de palavras por frase constitui um bom indicador da complexidade das frases, na medida em que frases mais longas proporcionam uma oportunidade aos autores de construir frases complexas. No entanto, dever-se-á ter em conta que a clareza do resumo é igualmente afectada pela complexidade do conteúdo que retrata, pelo que nem sempre frases mais complexas significam uma falta de qualidade na elaboração do resumo por parte dos indexadores.

Os serviços em análise apresentam uma média de 22.7 palavras por frase (Cf. Tabela 57), o que revela uma complexidade significativa das frases produzidas.

Os dados revelam grande homogeneidade (coeficiente de variação = 6%) e não se verifica a hipótese da maior ou menor complexidade da frase estar relacionada com as especificidades das áreas temáticas abrangidas por cada serviço, já que existe uma verdadeira pulverização dos serviços na escala de valores.

Evitar a utilização da voz passiva constitui um dos requisitos importantes para a legibilidade de qualquer texto. Mas, como afirma Wheatley, “as construções na voz passiva são um sintoma natural de escrever na terceira pessoa e no passado, que são comuns nas escritas técnicas, científicas e médicas, e são provavelmente confortavelmente familiares aos utilizadores alvo” (WHEATLEY:1997, 31).

Os dados recolhidos verificam, em certa medida, a hipótese de uma maior utilização da voz passiva nas áreas científicas das ciências exactas e da medicina (Cf. Tabela 58)

No entanto, na generalidade, a percentagem de utilização da voz passiva é alta, com uma média de 21.5%. Existe, no entanto, pouca homogeneidade entre estes valores (coeficiente de variação = 22%).

A aplicação dos testes Flesh Reading Ease e Flesch Kincaid Grade Level é reveladora das inconsistências já identificadas no estudo de Wheatley (WHEATLEY:1997). Enquanto que os resultados do primeiro apresentam variações dentro de cada serviço e entre os serviços, o segundo mantém invariavelmente o valor 12, independente de um valor mais alto ou mais baixo no outro. De qualquer forma, é possível analisar os resultados considerando-os como valores não absolutos, mas com carácter de indicador relativo de uma dada realidade.

A tendência geral é para um baixo nível de legibilidade em todos os serviços, com uma média de 20 pontos (Cf. Tabela 59).

Os dois grandes pólos são HUMBUL, com o maior nível de legibilidade e PSIGATE com o mais baixo. Os valores do meio da tabela, são de tal forma próximos, que tendo em conta as deficiências deste tipo de teste, não é possível atribuir-lhes um significado discriminador.

A correlação entre os resultados da legibilidade do teste Flesh Reading Ease (FRE) e os outros indicadores até aqui analisados permite ultrapassar a análise descritiva realizada até ao momento.

A primeira hipótese a verificar é a que diz respeito à relação entre a legibilidade e a extensão do resumo considerando que a extensão do resumo tem um impacto negativo sobre a legibilidade, ou seja, quanto maior for o resumo menos legível se torna.

Ordenando os dados da extensão do resumo de acordo com os resultados obtidos no teste FRE, tal hipótese não se verifica (Cf. Tabela 60). Ao contrário, se tivermos em atenção que o serviço HUMBUL, que apresenta a maior extensão média dos resumos é o que está posicionado em primeiro lugar no teste de legibilidade FRE, a eventual relação entre estas duas variáveis parece ser exactamente contrária à hipótese inicial. De facto, existe uma correlação positiva forte ( $r = +0.89$ ) entre as duas variáveis, o que significa que o aumento da variável extensão conduzirá ao aumento da variável legibilidade. A importância desta correlação pode ser melhor avaliada se atendermos ao resultado do coeficiente de determinação (coeficiente de determinação = 79.21%), cujo resultado permite entender que a legibilidade é condicionada em grande medida pela extensão do resumo. Este resultado contraria alguns estudos que apontam exactamente para um correlação de sentido contrário (WHEATLEY:1997).

A segunda hipótese a verificar encontra-se no âmbito da relação entre a legibilidade (teste FRE) e a extensão da frase, considerando que frases menos extensas aumentam a legibilidade.

Ordenando os dados da extensão da frase de acordo com os resultados obtidos no teste FRE, verifica-se não existir uma relação imediata entre as duas variáveis (Cf. Tabela 61), dado que as frases mais extensas ocorrem nos resumos mais e menos legíveis.

De facto, os resultados do cálculo de coeficiente de correlação ( $r = +0.30$ ) indicam uma correlação muito fraca, o que pode ter origem na interacção entre extensão da frase e extensão do resumo.

A terceira hipótese é formulada no domínio da possível relação entre a complexidade das palavras (em número de caracteres por palavra) e a legibilidade, considerando que palavras menos complexas influenciam decisivamente uma maior legibilidade do resumo.

Ordenando os dados relativos à complexidade da frase de acordo com os resultados obtidos no teste FRE, verifica-se a hipótese sugerida (Cf. Tabela 62), na medida em que os resumos mais legíveis são os que apresentam um valor médio menor para os número de caracteres por palavra, sendo a inversa igualmente verdadeira.

A correlação entre as duas variáveis é muito forte ( $r = -0.89$ ), significando que uma diminuição do valor da variável complexidade da frase fará aumentar a legibilidade do resumo. Assim, pode concluir-se que grande parte da legibilidade pode ser explicada pela maior ou menor complexidade da frase (coeficiente de determinação = 79.21%).

A hipótese da existência de uma correlação entre a legibilidade e o número de frases por parágrafo não se verifica (coeficiente de correlação = -0.039).

A análise da legibilidade das descrições produzidas pelos serviços objecto deste estudo permite traçar um perfil dos resumos: textos não estruturados com uma extensão média (102 palavras), frases com alguma complexidade (22.7 palavras por frase) mas utilizando palavras pouco complexas (5.5 caracteres por palavra), grande utilização da voz passiva (21.5% em média), reduzido nível de legibilidade (20 pontos em 100). As variáveis extensão do resumo (coeficiente de determinação = 79.21%) e complexidade das palavras (coeficiente de determinação = 79.21%) são as que mais directamente influenciam a legibilidade, parecendo apontar-se para que resumos mais extensos, mas mantendo ou mesmo diminuindo a complexidade das palavras utilizadas assegurariam uma maior legibilidade.

## Cap. V – Conclusões

*Subject Gateways* apresentam uma proposta de acesso a recursos *Web* para a comunidade científica que se distingue de outros serviços de acesso à *Web*, essencialmente pelas seguintes características: selecção dos recursos, com a qual se pretende assegurar sobretudo informação de qualidade adequada ao público-alvo; metainformação de qualidade baseada em linguagens de metadados estruturadas; opções de pesquisa da metainformação quer por palavra-chave, quer através da organização de estruturas temáticas que permitem o *browsing*; uma política de gestão da colecção que visa assegurar a continuidade do acesso aos recursos e o seu crescimento.

Este tipo de serviços aplica um modelo de organização da informação baseada em linguagens de metadados relativamente ricas de conteúdo, com uma clara predominância para os atributos que representam o assunto, com um recurso maioritário a vocabulários controlados. Estas linguagens representam uma alternativa, por um lado, à complexidade da descrição bibliográfica tradicional (modelo ISBD) e por outro à simplicidade da maioria das propostas mais recentes de metadados para objectos digitais.

Um outro aspecto particularmente relevante neste domínio é a aposta em estruturas temáticas para *browsing* das descrições, baseadas maioritariamente em linguagens de assunto já existentes e vindas do paradigma biblioteconómico tradicional. Em relação à pesquisa por palavra, esta forma de acesso é claramente pouco privilegiada, já que em muitos casos se verifica um reduzido aproveitamento da metainformação para basear funcionalidades de pesquisa, por exemplo, a inexistência de pesquisa por atributos específicos como o autor ou o título.

Considerando o público-alvo e o facto de que se pretende construir uma colecção seleccionada de recursos, estando, portanto, ausente qualquer preocupação de recuperar toda a *Web* mesmo nas áreas temáticas específicas, os objectivos que se pretendem atingir são aqueles que caracterizam qualquer sistema de informação de tipo bibliográfico, a saber: encontrar uma entidade singular, encontrar conjuntos de entidades de acordo com pontos de acesso mais relevantes (particularmente títulos, autores e assuntos), identificar uma entidade, seleccionar uma entidade apropriada à satisfação das necessidades do utilizador, obter acesso à entidade seleccionada e, finalmente, navegar no universo das descrições dessas entidades explorando diversos tipos de relações possíveis.

A concretização destes objectivos implica uma descrição mais complexa e estruturada, exigindo conhecimento profissional e custo relativamente elevado. A opção por este tipo de abordagem aplicada a toda a *Web* está, pelas razões indicadas, fora de questão. Mas, neste caso, revela-se pertinente, já que se trata, poderíamos designar, pelo melhor da *Web* para um público exigente.

Os mais antigos *Gateways* que actualmente compõem a RDN, objecto de análise neste trabalho, nasceram exactamente como uma solução possível para promover o acesso à *Web* de qualidade por parte da comunidade universitária inglesa e constituíram iniciativas de bibliotecas, muitas vezes em conjugação com organizações especializadas nas áreas temáticas objecto de inclusão. Estes serviços foram analisados a partir de um conjunto de critérios construídos com base num quadro teórico específico que considera as características e objectivos de um sistema de informação bibliográfico por referência às expectativas do público-alvo. Os critérios assim definidos foram aplicados, quer às linguagens de metadados de cada serviço, quer a uma amostra significativa das descrições produzidas, quer ainda às estruturas de *browsing*, permitindo assim uma dupla avaliação, do ponto de vista conceptual e do ponto de vista aplicacional.

Os recursos *Web* seleccionados por estes serviços são organizados tem por base linguagens de metadados específicas. Estas linguagens baseiam-se, do ponto de vista conceptual, nas propostas do *DESIRE information gateway handbook*, que privilegiam a descrição da **Obra** enquanto conteúdo intelectual com uma relativa desvalorização do objecto. Os atributos claramente privilegiados para a descrição desta entidade são os títulos e o assunto. As recomendações relativamente ao vocabulário utilizado apontam claramente no sentido do controlo vocabular em todas os atributos onde tal se revele mais importante, nomeadamente assuntos, nomes, datas, tipo de recurso. Do ponto de vista prático, a maioria das regras de catalogação dos *Gateways* RDN baseiam-se nas que foram definidas no âmbito do projecto ROADS, que prevêm originalmente um muito maior número de atributos da informação que os sugeridos pelo manual DESIRE, nomeadamente no que respeita ao **Documento** enquanto objecto “físico”, mas também em relação aos atributos da **Obra** com os novos atributos Versão, Discussão, Comentários do catalogador, Citação, Fonte e a distinção de 3 categorias independentes para o atributo palavras-chave (nomes, organizações e assuntos). No caso dos responsáveis pelo conteúdo e pela edição verifica-se também uma concretização que está relativamente longe do sugerido em DESIRE, quer pela consideração de uma extensa variedade de funções, quer ainda pelos atributos definidos para cada um dos tipos de responsabilidade. Posteriormente, o projecto ROADS criou um *template* para as descrições de recursos exclusivamente *Web* pensado para o contexto RDN e para se aproximar de Dublin Core. A análise das regras de catalogação dos serviços permitiu verificar que este *template* não foi aplicado.

As regras de catalogação seguidas pelos vários *Gateways* apresentam grandes semelhanças estruturais, em virtude da aplicação do software ROADS, exceptuando o caso de HUMBUL. As dissemelhanças verificam-se em dois domínios, sempre no sentido da redução do número de atributos: no caso dos atributos relativos à responsabilidade e nos atributos do **Documento**. Em relação às propostas de DESIRE todos os elementos sugeridos estão contemplados, revelando-se uma tendência por parte de todos os serviços para o seu desenvolvimento. Mesmo assim, cada uma destas linguagens prevê um extenso número de atributos, na grande maioria para a descrição da **Obra**, sendo atributos maioritariamente coincidentes (verificam-se, pelo menos, em três dos serviços analisados) os seguintes: Título próprio, Complemento de título, Título paralelo, Autor pessoa física, Descrição em texto livre, Palavras-chave, Classificação, Língua, Tipo de recurso. No caso de HUMBUL, prevêm-se outro tipo de atributos que permitem uma melhor concretização dos objectivos, nomeadamente: cobertura (geográfica e temporal), audiência e relação.

Exceptuando o caso dos assuntos, o nível de controlo vocabular é relativamente reduzido nestas linguagens e é apenas previsto para os casos das datas, língua, tipologia do recurso, mesmo assim nem em todos os casos.

Em termos de semântica, e mais uma vez excluindo o caso dos assuntos, são muito raros os dispositivos de controlo semântico. O mesmo se aplica para a sintaxe.

As linguagens do **Documento** preocupam-se sobretudo com os atributos localização e requisitos / condições de acesso (incluindo neste conjunto os direitos de autor) e muito menos com os atributos ligados à fisicalidade e à publicação.

As linguagens assim construídas revelam, sobretudo, potencialidade para concretizar os objectivos encontrar uma entidade singular, localizar todos os documentos com o mesmo assunto, prejudicado apenas pela ausência de qualquer regra de controlo de autoridade, obter acesso à entidade descrita e identificar e seleccionar uma entidade apropriada às necessidades do utilizador, neste último caso com um diferente nível de concretização entre os serviços. Destaque-se, no entanto, HUMBUL relativamente aos objectivos identificar e seleccionar, já que apresenta quase todos os atributos previstos como indicadores de concretização. Praticamente impossíveis de concretizar por todas as linguagens são os seguintes objectivos: localizar todos os documentos da mesma **Obra**, localizar todos os documentos da mesma versão, localizar todos os documentos de um mesmo autor e navegabilidade no universo bibliográfico.

A representação do conteúdo dos recursos, nos atributos Classificação e Palavras-chave, é maioritariamente realizada com o recurso a vocabulários controlados já utilizados no ambiente das bibliotecas. Utilizam-se quer vocabulários generalistas, quer vocabulários específicos com diferentes funções: no primeiro caso pretende-se dar uma visão generalista do conteúdo do serviço e proporcionar a reunião de um maior número de recursos sobre o mesmo tema, utilizando sobretudo classificações; no segundo uma aproximação mais específica e exaustiva, utilizando sobretudo tesouros. Os vocabulários seleccionados são apropriados, tendo em conta a diversidade mais ou menos abrangente das temáticas de cada serviço, na medida em que permitem representar com especificidade os conteúdos especializados.

A aplicação destas linguagens é realizada de forma maioritariamente consistente, pelo menos no que a amostra permitiu analisar, revelando-se os principais problemas no tratamento do atributo responsabilidade, autoral ou editorial, no que se exceptua o caso de HUMBUL. No entanto revela-se, em quase todas elas, um deficiente aproveitamento das potencialidades previstas nas linguagens para a concretização das expectativas do utilizador, traduzindo-se em: ausência nas descrições de atributos previstos na linguagem que permitiriam mais eficazmente concretizar os objectivos de identificar e seleccionar uma entidade (exceptua-se o caso de HUMBUL); a ausência de controlo de autoridade autor; a ausência de controlo de autoridade assunto (que produz discrepâncias assinaláveis se se pesquisar o mesmo assunto por palavra-chave ou na estrutura do *browsing*), a ausência de pesquisa por campos específicos (exceptuando para alguns campos o caso de SOSIG e EEVL). Assim, o objectivo localizar entidades singulares e conjuntos de entidades fica claramente comprometido, a não ser no caso dos assuntos, mas sobretudo por via do *browsing* e não da pesquisa por palavra.

A opção de pesquisa através do *browsing* de uma estrutura temática é a opção que recebe um tratamento mais cuidado de todos os serviços sem excepção. São utilizadas maioritariamente classificações, apenas BIOME utiliza também tesouros para esta função.

A implementação alfabética dos tesouros é claramente inapropriada, na medida em que estão ausentes quaisquer tipos de relações entre os termos que permitam ao utilizador seleccionar o mais apropriado antes de se decidir por qualquer hipótese.

A implementação das classificações é realizada de forma bastante flexível, evitando assim alguns inconvenientes de uma certa rigidez destas linguagens, como por exemplo, a representação num nível muito baixo da tabela de assuntos centrais para o serviço em causa. Esta flexibilidade de aplicação revela-se também na alta taxa de atribuição dos recursos aos assuntos verificada em quase todos os serviços, bem como na alteração terminológica de alguns descritores associados às notações de forma a adaptá-los melhor ao público-alvo. O modelo predominante de estruturação das classificações para navegação é o hierárquico. A aplicação deste modelo apresenta algumas virtualidades, a saber: permite uma visão abrangente do conteúdo do serviço, reunir todos os documentos de um mesmo assunto, definir estratégias de pesquisa várias de acordo com a situação de conhecimento de um dado tema por parte do utilizador, ou seja, possibilita ao utilizador começar ora por um assunto mais específico ora por um mais geral se não souber formular a questão na especificidade. Mas, verificam-se também aspectos menos positivos como sejam a ausência de outro tipo de relações entre os assuntos para além das hierárquicas, necessidade mais evidente nas estruturas mais complexas (exceptua-se o caso de SOSIG); a inexistência de qualquer relação entre a classificação utilizada e outras linguagens e, por último, a não utilização das notações, o que inviabiliza qualquer possibilidade de interoperabilidade futura.

Este modelo de implementação das classificações tem, na sua própria lógica, uma limitação estrutural: se a estrutura for muito generalista acumulam-se grandes conjuntos de descrições em cada assunto, tornando progressivamente mais difícil ao utilizador identificar os recursos apropriados; se se opta por uma estrutura mais complexa, existem duas prováveis desvantagens: aumenta grandemente a probabilidade de muitos assuntos serem representados por um reduzido número de descrições e aumenta também a exigência ao utilizador, no sentido de que a um aumento da complexidade da árvore temática terá que corresponder uma maior competência da sua leitura e descodificação por parte deste último. Consideramos que as soluções para esta limitação passam por: a) aplicar uma metodologia de gestão da árvore de assuntos que avalie sistematicamente a necessidade de expandir ou comprimir essa estrutura; b) conjugar o acesso à árvore por via da navegação das hiperligações com a pesquisa por palavra; c) desenvolver as relações entre os assuntos, mesmo que tal não seja previsto na linguagem original, por exemplo a atribuição de sinónimos.

A par do *browsing*, as descrições em texto livre apresentam-se também como um dos atributos aos quais é atribuída uma importância relevante. No entanto, a esta importância não corresponde efectivamente uma qualidade significativa na definição do seu conteúdo, primando pela ausência de parâmetros como tipo de resumo e sua estrutura claramente definidos e prestando-se mesmo a algumas confusões com outros atributos. Por outro lado, aspectos não presentes em outros atributos, até por não serem passíveis de grande normalização, como seria o caso da estrutura dos sítios *Web* ou de

uma avaliação das potencialidades de navegação estão completamente ausentes. A prática descritiva produziu um perfil de resumos com as seguintes características: textos não estruturados de média extensão, mas utilizando palavras pouco complexas, grande utilização da voz passiva, apresentando por isso um reduzido nível de legibilidade.

Num modelo descritivo que ficou a meio caminho entre a complexidade extrema e a simplicidade verificaram-se potencialidades várias, sobretudo no que diz respeito ao acesso por assunto e à qualidade das descrições que permite ao utilizador uma identificação e selecção efectiva do melhor recurso *Web* para responder às suas necessidades. A ausência de alguns atributos bem como o tratamento menos normalizado de outros produziu algumas inconsistências, mas sobretudo impossibilidade de corresponder a outras expectativas do exigente público-alvo. Neste último caso são de destacar o atributo responsabilidade, sobretudo responsabilidade intelectual, e o atributo relação. A pouca importância efectiva dada ao atributo autor é pouco entendível neste contexto, porque estamos perante informação seleccionada para especialistas e maioritariamente produzida por especialistas. A facilidade de estabelecer relações entre objectos de informação ou metainformação no contexto *Web* e por outro lado a multiplicidade de ligações possíveis entre as entidades informativas justificaria um aplicação mais consequente do atributo relação, passando desde logo pelo esforço de definição de uma taxonomia das relações para este ambiente.

Avaliando globalmente os serviços analisados podemos concluir que SOSIG apresenta a melhor performance no que diz respeito às possibilidades de acesso, com linguagens claramente apropriadas à colecção e ao público-alvo e com uma estrutura de *browsing* claramente eficaz, e HUMBUL uma melhor qualidade das descrições.

A utilização de linguagens de metadados ancoradas no modelo descritivo com origem na tradição biblioteconómica revela, assim, as suas potencialidades permitindo responder às principais expectativas dos utilizadores. Neste contexto, o recurso a linguagens de representação do assunto normalizadas, que foram objecto de processos de adaptação ao novo ambiente tecnológico, garante a consistência do ponto de acesso essencial, abrindo caminho para a construção de sistemas de acesso à informação que circula na *Web* qualitativamente mais eficazes.

## BIBLIOGRAFIA<sup>82</sup>

1. **ALCTS. CCS. SAC. SUBCOMMITTEE ON METADATA AND SUBJECT ANALYSIS** – *Subject data in the metadata record [Em linha]: recommendations and rationale*. 1999. [Consult. 10-01-2002] Disponível em WWW: <URL:<http://webserve.goust.edu/users/gddcasex/sac/metadata/report.html>>
2. **BECKER, Hans J. e outros** – *Evaluation report of partner subject gateways* [Em linha]. 2000. [Consult. 10-01-2002]. Disponível em WWW: <URL: <http://www.sub.uni-goetingen.de>>
3. **BECKER, Hans J. e outros** – *Survey of end user surveys*. 2000 (a) [Em linha]. [Consult. 04-02-2002]. Disponível em WWW: <URL: [http://www.renardus.org/about\\_us/deliverables/d6\\_6/D6\\_6\\_final.pdf](http://www.renardus.org/about_us/deliverables/d6_6/D6_6_final.pdf)>
4. **BERNERS-LEE, Tim** – *Metadata architecture* [Em linha]. 1997. [Consult. 04-02-2002]. Disponível em WWW: <URL: <http://www.w3.org/DesignIssues/Metadata.html>>
5. **BIOME collections policy**. [Consult. 07-02-2002]. Disponível em WWW: <URL: <http://biome.ac.uk/guidelines/collection>>
6. **BRADSHAW, Rebeca** – *Cataloguing rules for the ADAM database [Em linha]: a procedural manual*. [Consult. 18-02-2002]. Disponível em WWW: <URL: <http://www.adam.ac.uk/adam/reports/cat>>
7. **CAMPBELL, Debbie** – *Definitions for web-based services* [Em linha]. Canberra: National Library of Australia, 2001, actual. Fevereiro 2003. [Consult. 18-02-2003]. Disponível em WWW: <URL: <http://www.nla.gov.au/initiatives/sg/servicetypes.html>>
8. **CAMPOS, Fernanda Guedes de** – *Controlo de autoridade: novos contextos e soluções* [Em linha]. 2003. [Consult. 18-03-2004]. Disponível em WWW: <URL: <http://bnd.bn.pt/ed/jornadas2003/docs%5CControlodeAutoridade.doc>>
9. **CASE, Donald Owen** – *Looking for information: a survey of research on information seeking needs and behavior*. Londres: Academic Press, 2002
10. **CAUSTON, Laurie, e outros** – *Metadata watch report #8. Standards framework #4* [Em linha]. 2002. [Consult. 12-04-2002]. Disponível em WWW: <URL: [http://www.schemas-forum.org/metadata\\_watch/d29/d29.pdf](http://www.schemas-forum.org/metadata_watch/d29/d29.pdf)>
11. **CHAN, Lois Mai** – *Exploiting LCSH, LCC and DDC to retrieve networked resources* [Em linha]: *issues and challenges*. 2000. [Consult. 15-04-2002]. Disponível em WWW: <URL: [http://www.loc.gov/catdir/bibcontrol/chan\\_paper.html](http://www.loc.gov/catdir/bibcontrol/chan_paper.html)>
12. **CHAPMAN, Ann; DAY, Michael; HIOM, Debra** – *Metadata: cataloguing practice and internet subject-based information gateways. Ariadne* [Em linha]. Nº 18 (Dezembro 1998). [Consult. 16-08-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue18/metadata/>>
13. **CIOLEK, T. Matthew** – *The six quests for the electronic grail: current approaches to information quality in WWW resources. Review informatique et statistique dans les sciences humaines* [Em linha]. Nº1-4 (1996). [Consult. 20-

---

<sup>82</sup> Esta bibliografia foi construída de acordo com a NP 405-4:2001 – Informação e Documentação – referências bibliográficas - Parte 4: documentos electrónicos

- 05-2003]. Disponível em WWW: <URL: <http://www.ciolek.com/PAPERS/QUEST/QuestMain.html>>
14. **CLARK**, Judith – Subject portals. *Ariadne* [Em linha]. Nº 29 (Outubro 2001). [Consult. 24-06-2003]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue29/clark/>>
  15. **CLEVELAND**, Donald B.; **CLEVELAND**, Ana D. – *Introduction to indexing and abstracting*. Englewood: Libraries Unlimited, 2001
  16. **DAY**, Michael – *ROADS cataloguing guidelines* [Em linha]: *revised version* (1.0). [Consult. 17-05-2003]. Disponível em WWW: <URL: <http://www.ukoln.ac.uk/metadata/roads/cataloguing/cataloguing-rules.html>>
  17. **DAY**, Michael; **CLIFF**, Peter – *RDN cataloguing guidelines* (versão 1.1) [Em linha]. 2000, actual. Outubro 2003. [Consult. 26-10-2003]. Disponível em WWW: <URL: <http://www.rdn.ac.uk/publication/cat-guide>>
  18. *DCMI metadata terms* [Em linha]. 2004. [Consult. 23-07-2004]. Disponível em WWW: <URL: <http://www.dublincore.org/documents/dcmi-terms>>
  19. **DEMPSEY**, Lorcan; **HEERY**, Rachel - *A review of metadata* [Em linha]: *a survey of current resource description formats*. 1997. [Consult. 06-08-2002]. Disponível em WWW: <URL: <http://www.surfnet.nl/innovatie/desire1/deliver/WP3/D32-1.html>>
  20. **DEMPSEY**, Lorcan; **GARDNER**, Tracy; **DAY**, Michael – International information gateway collaboration: report of the first Imesh framework workshop. *D-Lib Magazine* [Em linha]. Vol. 5, nº 12 (Dezembro 1999). [Consult. 01-09-2003]. Disponível em WWW: <URL: <http://www.dlib.org/dlib/december99/12dempsey.html>>
  21. **DEMPSEY**, Lorcan – The subject gateway [Em linha]: experiences and issues based on the emergence of the Resource Discovery Network. *Online Information Review* [Em linha]. Nº 24: 1 (2000). [Consult. 12-08-2003]. Disponível em WWW: <URL: <http://www.rdn.ac.uk/publications/ior-2000-02-dempsey/>>
  22. **DESIRE INFORMATION GATEWAYS HANBOOK**. Actual. Abril 2000. [Consult. 13-03-2002]. Disponível em WWW: <URL: <http://www.desire.org/hanbook>>
  23. **DILLON**, Martin - *Metadata for web resources* [Em linha]: *how metadata works on the Web*. 2001. [Consult. 09-05-2002]. Disponível em WWW: <URL: <http://leweb.loc.gov/catdir/bibcontrol/dillon-paper.html>>
  24. **DOCHARTAIGH**, Niall O. – *The Internet research handbook*. Londres: Sage, 2002
  25. *Dublin Core metadata element set* [Em linha]: *version 1.3: reference description*. 2003. [Consult. 15-09-2003]. Disponível em WWW: <URL: <http://www.dublincore.org/document/dces>>
  26. **EDWARDS**, Alun; **FRASER**, Michael – *Collection development policy* [Em linha]. 2001. [Consult. 12-08-2002]. Disponível em WWW: <URL: <http://www.humbul.ac.uk/about/colliden.html>>
  27. **EDWARDS**, Alun; **FRASER**, Michael – *Collection development policy* [Em linha]. 2001, actual. Setembro 2003. [Consult. 26-10-2003]. Disponível em WWW: <URL: <http://www.humbul.ac.uk/about/colldev.html>>
  28. *EEVL collection development policy*. [Consult. 07-11-2002]. Disponível em WWW: <URL: <http://www.eevl.ac.uk/cdp.htm>>
  29. **FOLLET**, Brian – *Follet report*. 1993. [Consult. 09-06-2003]. Disponível em WWW: <URL: <http://www.ukoln.ac.uk/services/papers/follet/report/ch1.html>>
  30. **FRASER**, Michael; **EDWARDS**, Alun – *Describing and cataloguing resources* (HUMBUL) [Em linha]. 2001. [Consult. 1-02-2002]. Disponível em WWW: <URL: <http://www.humbul.ac.uk/about/catalogue.html>>

31. **GARCÍA MATEU, M. A Jesús; SÀNCHEZ SANZ, T.** – *Internet subject gateways knowledge organization system evaluation* [Em linha]. 2001. [Consult. 22-02-2002]. Disponível em WWW: <URL: <http://pacs.única.it/alghero2001/proceedings/056.htm>>
32. **GORMAN, Michael** – Metadata: hype and glory. In: *Cataloging the Web: metadata, AACR, and Marc21*. Londres: Scarecrow Press, 2002, pp. 235-249
33. **A guide to ROADS**.1997. [Em linha]. [Consult. 27-03-2002]. Disponível em WWW: <URL: [www.ilrt.bris.ac.uk/roads/papers/guide.html](http://www.ilrt.bris.ac.uk/roads/papers/guide.html)>
34. **GRAY, Lisa** – *Cataloguing rules for BIOME service* [Em linha]: *a procedural manual*. [Consult. 13-06-2002]. Disponível em WWW: <URL: <http://biome.ac.uk/guidelines/cat/>> Actualmente indisponível.
35. **HAYNES, David, e outros** – *Comparative evaluation of the subject based gateways approach to providing access to network resources* [Em linha]: *a report to JISC under the eLib supporting studies programme*. 1998. [Consult. 07-10-2002]. Disponível em WWW: <URL: <http://www.ukoln.ac.uk/services/elib/>>
36. **HIOM, Debra** – *Sosig cataloguing rules* [Em linha]. 1999. [Consult. 07-10-2002]. Disponível em WWW: <URL: [http://www.sosig.ac.uk/about\\_us/docs/catrul.html](http://www.sosig.ac.uk/about_us/docs/catrul.html)>
37. **HJORLAND, Birger** – *Information seeking and subject representation: an activity-theoretical approach to information science*. Londres: Greenwood Press, 1997
38. **HOOPER, Tracey** – Planet\_Sosig: the new Sosig interface. *Ariadne* [Em linha]. Nº 10 (Julho 1997). [Consult. 7-10-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue10/planet-sosig/>>
39. **HOWARTH, Lynne C.** – *Metadata schemas for Subject Gateways* [Em linha]. 2003. [Consult. 25-11-2003]. Disponível em WWW: <URL: <http://www.ifla.org/IV/ifla69/papers/053e-Howarth.pdf>>
40. **HSIEH-YEE** – Organizing internet resources: teaching cataloging standards and beyond. *OCLC systems & services*. Vol. 16, nº3, 2000, pp.130-143
41. **HUXLEY, Lesly** – Planet Sosig has a new satellite: the Social Science\_Research Grapevine. *Ariadne* [Em linha]. Nº 15 (1998). [Consult. 07-10-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue15/>>
42. **IFLA Study Group on the Functional Requirements for Bibliographic Records** – *Functional requirements for bibliographic records*. Munique: K.G. Saur, 1998. [Consult. 10-01-2002]. Disponível em WWW <URL: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>>
43. **International standard bibliographic description** [Em linha]: *electronic resources*. [Consult. 6-08-2002]. Disponível em WWW: <URL: <http://www.ifla.org/VII/s13/pubs/isbd.htm>>
44. **Introduction to Dewey Decimal classification.** [Consult. 12-11-2002]. Disponível em WWW: <URL:<http://www.oclc.org/dewey/versions/ddc22print/info.pdf>>
45. **JENNINGS, Simon** – *RDN collections development framework* (versão 1.2) [Em linha]. 2002. [Consult. 12-11-2002]. Disponível em WWW: <URL: <http://www.rdn.ac.uk/publications/collections/cdframework3.doc>>
46. **JOINT STEERING COMMITTEE for revision of Anglo-American cataloguing rules** – *A brief history of AACR*. [Em linha]. Actual Julho 2004. [Consult. 07-07-2004]. Disponível em WWW: <URL: <http://www.nlc-bnc.ca/jsc/history.html>>

47. **KERR**, Linda – Casting an EEVL eye over engineering resources. *Ariadne* [Em linha]. Nº 9 (Maio 1997). [Consult. 08-10-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue9/eevl/>>
48. **KITCHEN**, Justine – New millennium, new Sosig. *Ariadne* [Em linha]. Nº 23 ( Março 2000). [Consult.8-10-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue23/planet-sosig/intro.html>>
49. **KIRRIEMUIR**, John – *A brief survey of quality resource discovery systems* [Em linha]. 1999. [Consult. 09-10-2002]. Disponível em WWW: <URL: <http://www.rdn.ac.uk/publications/studies/survey/>>
50. **KOCH**, Traugott, e outros – *Specification for resource description methods part 3* [Em linha]: *the role of classification schemes in Internet resource description and discovery*. [Consult. 09-10-2002]. Disponível em WWW: <URL: [http://www.ub.lu/desire/radar/reports/D3.2.3/class\\_v10.html](http://www.ub.lu/desire/radar/reports/D3.2.3/class_v10.html)>
51. **KOCH**, Traugott – Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review* [Em linha]. Vol. 24:1 (Fevereiro 2000). [Consult. 12-03-2002]. Disponível em WWW: <URL: <http://www.emeraldinsight.com/oir.htm>>
52. **LANCASTER**, F. W. – Indexação e resumos: teoria e prática. Brasília: Briquete de Lemos, imp. 1993
53. **LAGOZE**, Carl - Keeping Dublin Core simple: cross-domain discovery or resource description?. *D-Lib Magazine* [Em linha]. Volume 7, nº 1 (Janeiro 2001). [Consult. 12-03-2002] Disponível em WWW: <URL: <http://www.dlib.org/dlib/january01/lagoze/01lagoze.html>>
54. **LAW**, Derek and **DEMPSEY**, Lorcan (2000) – A policy context: eLib and the emergence of the subject gateways. *Ariadne* [Em linha]. Nº 25 ( Setembro 2000). [Consult. 12-03-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue25/subject-gateways/>>
55. **LYNCH**, Clifford – Future developments in metadata and their role in access to networked information. In: *Cataloguing the web: metadata, AACR, and Marc21*. Londres: Scarecrow Press, 2002, pp. 284-298
56. **MACLEOD**, Rody – EEVL: new profile, new office and coming soon: a new service. *Ariadne* [Em linha]. Nº 26 (Janeiro 2001). [Consult. 22-04-2002] Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue26/eevl/intro.html>>
57. **MACLEOD**, Rody; **KERR**, Linda – EEVL update. *Ariadne* [Em linha]. Nº 28 (Junho 2001). [Consult. 22-04-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue28/eevl/intro.html>>
58. **McCALLUM**, Sally - *Extending MARC for bibliographic control in the Web environment* [Em linha]: *challenges and alternatives*. 2000. [Consult.18-05-2002]. Disponível em WWW: <URL: [http://leweb.loc.gov/catdir/bibcontrol/mccallum\\_paper.html](http://leweb.loc.gov/catdir/bibcontrol/mccallum_paper.html)>
59. **MARTIN**, Ruth – Turning Gateways into Portals. *Library and Information Update* [Em linha]. (Junho 2003). [Consult.28-10-2003]. Disponível em WWW: <URL: <http://www.cilip.org.uk/update/issues/jun03/article3june.html>>
60. **MEDEIROS**, Norman – Driving with eyes closed: the perils of traditional catalogs and cataloguing in the Internet age. *Library computing*. Vol. 18, nº4 (1999), pp.300-315
61. **MÉNDEZ RODRÍGUEZ**, Eva – *Metadatos y recuperación de información: estándares, problemas y aplicabilidade en bibliotecas digitales*. Gijón: Ediciones Trea, 2002
62. **METCALFE**, Randy – My Humbul: Humbul gets personal. *Ariadne* [Em linha]. Nº 30 (Dezembro 2001). [Consult. 25-09-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue30/myhumbul/>>

63. **MILLER**, Libby – *ImeshTK* [Em linha] : *subject gateway review literature review*. 2000. [Consult. 18-03-2002]. Disponível em WWW: <URL: <http://www.imesh.org/>>
64. **MILLER**, Ken – *Meaningful relationships* [Em linha]. 1997. [Consult. 28-09-2003]. Disponível em WWW: <URL: <http://www.iassistdata.org/publications/iq/iq21/iqvol214miller.pdf>>
65. **MILSTEAD**, Jessica; Feldman, Susan - *Metadata: projects and standards. Online* [Em linha]. (Janeiro/Fevereiro 1999). [Consult. 09-10-2002]. Disponível em WWW: <URL: <http://www.onlineinc.com/onlinemag/OL1999/milstead1.html#projects/>>
66. **NAVES**, Madalena Martins Lopes – *Aspectos conceituais do browsing na recuperação da informação*. [Consult. 13-04-2003]. Disponível em WWW:<URL: <http://www.ibict.br/cienciadainformacao/include/getdoc/php?id=6618&article=348&mode=pdf>>
67. **NICHOLSON**, Scott – A proposal categorization and nomenclature for web search tools. In: *Internet search and indexing: the subject approach*. Nova Iorque: Haworth Press, 2000, pp. 9-28
68. **OLIVEIRA**, José Manuel Paquete de, e outros – *The Internet as a construction tool for a participated citizenship*. 2000. [Consult. 20-12-2001]. Actualmente indisponível
69. **OLSON**, Nancy – *Cataloguing Internet resources: a manual and practical guide*. Dublin, Ohio: OCLC On-Line Computer Library Center, 1995, actual. 22 Jun. 1999. [Consult. 12-01-2002]. Disponível em WWW: <URL: <http://www.oclc.org/oclc/man/9256cat/toc.htm>>
70. **PINTO**, Maria; LANCASTER, F.W. – Abstracts and abstracting in knowledge discovery. *Library trends*. Vol. 48, nº 1.
71. **PITSCHMAN**, Louis A. – Building sustainable collections of free third-party web resources. 2001. [Consult. 15-04-2003]. Disponível em WWW: <URL: <http://www.clir.org/pubs/reports/pub98/pub98.pdf>>
72. **PSIGATE** – *Procedure for adding resources to the Psigate internet resource catalogue* (V. 8) [Em linha]. Setembro 2003. [Consult. 12-11-2003]. Disponível em WWW: <URL: <http://www.psigate.ac.uk/training2.doc>>
73. **PSIGATE** – *Introduction and procedure for locating and selecting resources to the Psigate internet resource catalogue (IRC)* (V. 10). [Em linha]. Setembro 2003. [Consult.12-11-2003]. Disponível em WWW: <URL: <http://www.psigate.ac.uk/training1.doc>>
74. **PSIGATE** – *Resource scope* [Em linha]: *full version* (V.10). Setembro 2003. [Consult. 12-11-2003]. Disponível em WWW: <URL: <http://www.psigate.ac.uk/content2a.doc>>
75. **PSIGATE** – *Resource evaluation criteria*(V. 7). [Em linha]. Setembro 2003. [Consult. 12-11-2003]. Disponível em WWW: <URL: <http://www.psigate.ac.uk/content3.doc>>
76. **PSIGATE** – *Cataloguing rules* (V.8). [Em linha]. Setembro 2003. [Consult. 12-11-2003] Disponível em WWW: <URL: <http://www.ac.uk/content4.doc>>
77. **Resource evaluation for BIOME**. [Consult. 23-07-2003]. Disponível em WWW: <URL: <http://biome.ac.uk/guidelines/eval>>
78. **ROWLEY**, Jennifer; FARROW, John – *Organizing knowledge: an introduction to managing access to information*. Londres: Grower, 2000
79. **SCHWARTZ**, Candy – *Sorting out the web: approaches to subject access*. London: Alex Publishing, 2001

80. **SELBERG**, Erik Warren - *Towards comprehensive Web search* [Em linha]. Washington: University of Washington, 1999 (tese de doutoramento defendida na Universidade de Washington para a obtenção de PhD em Computer Science & Engineering). [Consult. 24-03-2002]. Disponível em WWW: <URL: [http://citeseer.ist.psu.edu/cache/papers/cs/15516/httpzSzzSzwww.mri.mq.edu.auzSz~einatzSzweb\\_irzSzselberg\\_thesis.pdf/selberg99towards.pdf](http://citeseer.ist.psu.edu/cache/papers/cs/15516/httpzSzzSzwww.mri.mq.edu.auzSz~einatzSzweb_irzSzselberg_thesis.pdf/selberg99towards.pdf)>
81. **SLADEN**, Catherine; **SPENCE**, Mary – Hand picked for quality: a reflection on Biz/ed. *Library Consortium Management*. Vol. 2, nº 2, (2000), pp. 35-43
82. **SOON**, Chan – *Interdisciplinary perspectives on the quality of abstracts for information retrieval*. [Em linha]. 2002. [Consult.02-05-2002]. Disponível em WWW: <URL: <http://www.rdn.ac.uk/abstracts/abstracts.pdf>>
83. **SOSIG** – *Selection criteria*. [Consult. 09-03-2003]. Disponível em WWW: <URL: <http://www.sosig.ac.uk/desire/ecrit.html>>
84. **SOSIG** – *Scope policy*. [Consult. 09-03-2003]. Disponível em WWW: <URL: <http://www.sosig.ac.uk/desire/escope.html>>
85. **SOSIG** – *Collection management policy*. [Consult. 09-03-2003]. Disponível em WWW: <URL: [http://www.sosig.ac.uk/about\\_us/collect.html](http://www.sosig.ac.uk/about_us/collect.html)>
86. **SVENONIUS**, Elaine – *The intellectual foundation of information organization*. Cambridge: MIT press, 2001
87. **TAYLOR**, Arlene G. – *The organization of information*. Nova Iorque: Libraries Unlimited, 1999
88. **THOMAS**, Alan R., **SHEARER**, James (ed.) – *Internet searching and indexing: the subject approach*. Nova Iorque: Haworth Press, 2000
89. **TILLET**, Barbara B. - *Authority control on the Web* [Em linha]. 2001. [Consult. 26-08-2002]. Disponível em WWW: <URL: [http://www.loc.gov/catdir/bibcontrol/tillet\\_paper.html](http://www.loc.gov/catdir/bibcontrol/tillet_paper.html)>
90. **VIZINE-GOETZ**, Diane – *Using library classification schemes for Internet resources* [Em linha]. [Consult.24-08-2002]. Disponível WWW: <URL: <http://www.oclc.org>>
91. **VIZINE-GOETZ**, Diane – Classification schemes for Internet resources revisited. In: *High-level subject access tools and techniques in Internet cataloging*. Nova Iorque: The Haworth Information Press, 2002, pp. 5-18
92. **WHEATLEY**, Alan; **ARMSTRONG**, C.J.U. – *A survey of the content and characteristics of electronic abstracts* [Em linha]. 1997. [Consult. 18-04-2002]. Disponível em WWW: <URL: <http://www.ukoln.ac.uk/dlis/models/studies/abstracts/abstracts.doc>>
93. **WILSON**, T.D. – Human information behaviour. *Informing Science* [Em linha]. Vol.3, nº2 (2000). [Consult. 07-08-2002]. Disponível em WWW: <URL: <http://inform.nu/articles/vol3/v3nlp49-59.pdf>>
94. **WISEMAN**, Norman – International collaboration on subject based Internet Gateways. *D-Lib Magazine* [Em linha]. (Outubro 1998). [Consult. 07-04-2002]. Disponível em WWW: <URL: <http://www.blib.org/october98/10clips.html>>
95. **WOODCOCK**, Louise; **WHARAM**, Helen – Sosig expands into Europe. *Ariadne* [Em linha]. Nº 29 (2001). [Consult. 07-04-2002]. Disponível em WWW: <URL: <http://www.ariadne.ac.uk/issue29/planet-sosig/intro.html>>

## TABELAS

**Tabela 13 – Entidades e Atributos em SOSIG**

<b>Entidades</b>	<b>Atributos / Etiquetas</b>	<b>Definição</b>	<b>Utilização O= Obrigatório O'= Opcional</b>
<b>Obra</b>	Título	O título próprio	O
	Título alternativo	Uma alternativa ao título, ou seja, o título em outra língua	O'
	Citação	O método preferido de citação deste recurso quando utilizado em outras obras	O'
	Categoria	Tipo de material	O
	Descrição	Descrição do recurso	O
	Palavras-chave	Palavras-chave apropriadas para descrever o conteúdo do recurso	O
	Língua	A língua em que o recurso está escrito	O
	Assunto	Ver entidade Assunto	O
	Autor	Ver entidade Autor	O'
	Direitos de autor	A declaração de direitos de autor	O'
	Fonte	Quando o recurso é uma cópia electrónica de uma versão em papel	O'
	Versão	O designativo da versão do recurso	O'
<b>Documento</b>	Política de acesso	Detalhes sobre qualquer restrição de acesso ao recurso	O'
	Tempos de acesso	Detalhes sobre o tempo em que o recurso está disponível	O'
	Tipo de caracteres	O tipo de caracteres do recurso	O'
	Preços de acesso	Detalhes sobre qualquer tipo de encargos com o recurso	O'
	Editor	Ver entidade Organização	O'
	Detentor do recurso	Ver entidade Organização	O'

	Administrador	Ver entidade Administrador do recurso	O'
	Patrocinador	Ver entidade Organização	O'
	Formato	Formato no qual o recurso está disponível	O'
	Requisitos técnicos	Quaisquer requisitos para a utilização do objecto	O'
	Registo	Detalhes sobre qualquer forma de registo que o recurso possa requerer	O'
	Tamanho	Tamanho do recurso em <i>bytes</i>	O'
	URI	O endereço Internet do recurso	O
	Autenticação	Detalhes sobre qualquer nome de utilizador ou <i>password</i> que o recurso possa requerer	O'
	Data de criação		O'
	Data da última revisão		O'
	Estado de publicação	Estado de publicação corrente do recurso (rascunho, publicado, etc.)	O'
	ISBN		O'
	ISSN		O'
Autor	Nome	Nome do indivíduo	O'
	Endereço postal do emprego	Endereço postal completo do indivíduo	O'
	País	País no qual o indivíduo trabalha	O'
	Telefone do emprego	Nº de telefone do emprego do indivíduo	O'
	Fax do emprego	Nº do fax do emprego do indivíduo	O'
	Título profissional	Título profissional do indivíduo	O'
	Departamento	Nome do departamento em que o indivíduo trabalha	O'
	Endereço electrónico	Endereço electrónico do indivíduo	O'
Administrador: responsável pela manutenção do recurso			
	Nome	Nome do indivíduo	O'
	Endereço postal do emprego	Endereço postal completo do indivíduo	O'
	País	País no qual o indivíduo trabalha	O'
	Telefone do emprego	Nº de telefone do emprego do indivíduo	O'
	Fax do emprego	Nº do fax do emprego do indivíduo	O'

	Título profissional	Título profissional do indivíduo	O'
	Departamento	Nome do departamento em que o indivíduo trabalha	O'
	Endereço electrónico	Endereço electrónico do indivíduo	O
Organização			
	Nome	Nome do indivíduo	O'
	Tipo	Tipo de organização	O'
	Endereço postal	Endereço postal completo da organização	O'
	País	País em que a organização está sediada	O'
	Telefone do emprego	Nº de telefone do emprego do indivíduo	O'
	Fax do emprego	Nº de fax do emprego do indivíduo	O'
	Endereço electrónico	Endereço electrónico da organização	O'
Assunto	Descritor	As notações da classificação para o recurso	O
	Esquema do descritor	O esquema de classificação utilizado para construir as secções de assunto	O (Automaticamente atribuído)

Fonte: SOSIG cataloguing rules

**Tabela 14 - Nível de consecução dos objectivos universais dos sistemas de informação bibliográfica em SOSIG**

<b>Objectivos</b>	<b>Indicadores</b>	<b>SOSIG Legenda: S = Verifica na totalidade N= Não Verifica na totalidade SN= Verifica não totalmente</b>
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título	S
	Atributo Título alternativo	S
	Atributo Versão	SN
	Atributo autor (atributos que permitem distinguir autores semelhantes)	S
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua	S
	Atributo Formato	S
	Atributo Requisitos de Acesso	S
	Atributo Descrição	S
	Atributo Assuntos	S
	Atributo Público-alvo	N
	Atributo Tipologia do recurso	S

	Atributo Data	S
	Atributo Versão	SN
	Atributo Cobertura	N
	Atributo Editor	S
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme	N
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: "É versão de" e "Tem versão"	N
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)	N
Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado	S
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)	N
Obter acesso à entidade descrita	Atributo URL / URN / URI	S
	Atributo ISBN / ISSN	S
Navegabilidade no universo bibliográfico	Atributo Relação	SN
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos	S

**Tabela 15 - Entidades e Atributos definidos em PSIGATE**

Entidade	Atributos / Etiquetas	Definição	Utilização (O= Obrigatório O'= Opcional)
<b>Obra</b>	Título	Título completo do recurso ou objecto	O
	Complemento de título	(sem definição específica)	O
	Título alternativo	Um título alternativo	O' (utilizado apenas para recursos em línguas que não o inglês)
	Palavras-Chave	Palavras-chave apropriadas para o recurso	O. Retiradas de tesouros apropriados
	Descrição	Descrição (isto é "resumo" no caso dos documentos) do recurso	O
	Categoria	Tipo do recurso, por exemplo, jornal electrónico, lista de distribuição	O
	Autor	Ver entidade Autor	O'
	Descritor de Assunto	Ver entidade Assunto	O'

	Direitos de autor	Declaração de direitos de autor. Qualquer informação adicional sobre a política de cópia pode ser incluída.	O'
	Língua	O nome da língua em que o recurso está escrito	O
<b>Documento</b>	Editor	Ver entidade Editor	O'
	Admin	Ver entidade Administrador	
	URI	Identificador uniforme do recurso; descrição do acesso ao recurso	O
Autor: a pessoa ou pessoas intelectualmente responsáveis pelo recurso	Nome	Nome do indivíduo	O'
	Endereço electrónico	Endereço electrónico do autor	O'
Editor: descrição / informação de contacto do editor do recurso	Nome	Nome do editor	O'
	Endereço electrónico	Endereço electrónico do editor	O'
Assunto	Descritor	Notação retirada de uma classificação	O'
	Esquema	Nome do esquema de classificação [DDC21]	O'
Administrador: a pessoa ou pessoas responsáveis pela manutenção e administração do recurso	AdminNome	A pessoa ou pessoas responsáveis pela manutenção e administração do recurso	O'
	Endereço electrónico	Endereço electrónico de AdminNome	O'

Fonte: Cataloguing rules

**Tabela 16 - Nível de consecução dos objectivos universais dos sistemas de informação bibliográfica em PSIGATE**

<b>Objectivos</b>	<b>Indicadores</b>	<b>PSIGATE Legenda: S = Verifica na totalidade N= Não Verifica na totalidade SN= Verifica não totalmente</b>
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título	S
	Atributo Título alternativo	S
	Atributo Versão	S
	Atributo autor (atributos que permitem distinguir autores semelhantes)	SN
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua	S
	Atributo Formato	N
	Atributo Requisitos de Acesso	N
	Atributo Descrição	S

	Atributo Assuntos	S
	Atributo Público-alvo	N
	Atributo Tipologia do recurso	S
	Atributo Data	S
	Atributo Versão	N
	Atributo Cobertura	N
	Atributo Editor	S
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme	N
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: "É versão de" e "Tem versão"	N
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)	N
Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado	S
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)	N
Obter acesso à entidade descrita	Atributo URL / URN / URI	S
Navegabilidade no universo bibliográfico	Atributo Relação	N
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos	S

**Tabela 17 - Entidades e Atributos definidos em BIOME**

Entidade	Atributos / Etiquetas	Definição	Utilização
<b>Obra</b>	Título	Título completo do recurso	
	Complemento de título	(sem definição específica)	
	Categoria	Tipo do recurso	
	Descrição	Descrição (ou seja, "resumo" no caso dos documentos) do recurso	
	Língua principal	O nome da língua em que o recurso está escrito	Só para línguas que não o inglês
	Fonte	Informação de acordo com a versão definitiva do recurso	
	Comentário	Comentários do catalogador	
	Autor	Ver entidade autor	
	Organização	Ver entidade organização	
	Audiência		Não utilizado actualmente

	Palavras-chave	Palavras-chave apropriadas para o recurso	De acordo com os tesouros usados por cada serviço
	Descritor de assunto	Notação retirada de uma classificação	De acordo com as classificações utilizadas pelos vários serviços
	Esquema do descritor de assunto	Nome do esquema utilizado	
	Direitos de Autor	Declaração dos direitos de autor. Qualquer informação adicional sobre a política de cópia	
<b>Documento</b>	URI Principal ( <i>Uniform Resource Identifier</i> )	Identificador uniforme do recurso; descrição do acesso ao recurso	
	Formato	Formatos em que o recurso está disponível	
	Data de criação		
	ISSN principal	Número internacional normalizado para publicações em série do recurso	
	ISBN principal	Número internacional normalizado para livros do recurso	
	Estatuto de publicação	Actual estatuto da publicação (rascunho, publicado, etc.)	
	Política de preços	Qualquer mecanismo de cobrança, incluindo estrutura de taxas se apropriado	
	URI <i>mirror</i>	Identificador uniforme do recurso; descrição do acesso alternativo ao recurso	
	Localização UK	O sítio é ou não baseado no UK	
Autor: a pessoa ou pessoas intelectualmente responsáveis pelo recurso	Hub	O serviço em que o recurso se encontra	
	Nome	Nome do indivíduo	
	Telefone do emprego	O número de telefone do emprego do indivíduo	
	Fax do emprego	Número de fax do indivíduo	
	Endereço postal do emprego	Endereço postal do indivíduo	
	País	País do indivíduo	
	Título da função	Título da função do indivíduo (se apropriado)	
	Departamento	Departamento ao qual o indivíduo pertence (se apropriado)	
	Endereço electrónico	Endereço electrónico do indivíduo	

Organização: a organização responsável pelo recurso	Hub	O serviço em que o recurso se encontra	
	Nome	Nome da organização	
	Endereço postal	Endereço postal da organização	
	Cidade	Cidade da organização	
	Estado	Estado, província ou país da organização	
	País	País da organização	
	Endereço electrónico	Endereço electrónico da organização	
	Telefone	Número de telefone da organização	
Fax	Número de fax da organização		

Fonte: Cataloguing rules for the Biome service: a procedural manual

**Tabela 18 - Nível de consecução dos objectivos universais dos sistemas de informação bibliográfica em BIOME**

<b>Objectivos</b>	<b>Indicadores</b>	<b>BIOME Legenda: S = Verifica na totalidade N= Não Verifica na totalidade SN= Verifica não totalmente</b>
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título	S
	Atributo Título alternativo	S
	Atributo Versão	N
	Atributo autor (atributos que permitem distinguir autores semelhantes)	S
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua	S
	Atributo Formato	S
	Atributo Requisitos de Acesso	SN
	Atributo Descrição	S
	Atributo Assuntos	S
	Atributo Público-alvo	N
	Atributo Tipologia do recurso	S
	Atributo Data	S
	Atributo Versão	N
	Atributo Cobertura	N
Atributo Editor	N	
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme	N
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: "É versão de" e "Tem versão"	N
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)	N

Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado	S
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)	N
Obter acesso à entidade descrita	Atributo URL / URN / URI	S
	Atributo ISBN / ISSN	N
Navegabilidade no universo bibliográfico	Atributo Relação	N
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos	S

**Tabela 19 - Entidades e atributos em EEVL**

Entidades	Atributos/Etiquetas	Definição	Utilização
<b>Obra</b>	Título	Sem informação	Sem informação
	Títulos alternativos	Sem informação	Sem informação
	Tipo de recurso	Sem informação	Sem informação
	Palavras-chave	Sem informação	Sem informação
	Classificação	Sem informação	Sem informação
	Descrição	Sem informação	Sem informação
	Comentários	Sem informação	Sem informação
	Autor (atributo retirado a partir da análise da amostra)	Sem informação	Sem informação
	Contacto- endereço electrónico	Sem informação	Sem informação
<b>Documento</b>	URL	Sem informação	Sem informação
	Localização	Sem informação	Sem informação
	ISBN/ISSN	Sem informação	Sem informação
	Autenticação	Sem informação	Sem informação
	Registo	Sem informação	Sem informação

**Tabela 20 - Nível de consecução dos objectivos universais dos sistemas de informação bibliográfica em EEVL**

Objectivos	Indicadores	EEVL <i>Legenda:</i> S = Verifica na totalidade N= Não Verifica na totalidade SN= Verifica não totalmente
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título	N
	Atributo Título alternativo	S
	Atributo Versão	N
	Atributo autor (atributos que permitem distinguir autores semelhantes)	SN
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua	N
	Atributo Formato	N
	Atributo Requisitos de Acesso	SN
	Atributo Descrição	S
	Atributo Assuntos	S
	Atributo Público-alvo	N

	Atributo Tipologia do recurso	S
	Atributo Data	N
	Atributo Versão	N
	Atributo Cobertura	N
	Atributo Editor	N
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme	N
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: "É versão de" e "Tem versão"	N
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)	N
Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado	S
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)	N
Obter acesso à entidade descrita	Atributo URL / URN / URI	S
	Atributo ISBN / ISSN	N
Navegabilidade no universo bibliográfico	Atributo Relação	N
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos	S

**Tabela 21- Entidades e Atributos em Humbul**

<b>Entidades</b>	<b>Atributos/Etiquetas</b>	<b>Definição</b>	<b>Utilização</b> Legenda: O= obrigatório O'= opcional
<b>Obra</b>	Título	Nome formal do recurso	O
	Complemento de título	(sem definição específica)	O'
	Título alternativo	Título abreviado, título traduzido ou qualquer outra forma de título diferente do título próprio.	O'
	Responsabilidade	Pessoas ou organizações responsáveis pela criação ou com contributo significativo para o recurso. Ver entidade responsabilidade	
	Descrição		O
	Audiência	Publico-alvo do recurso	O'
	Língua		O
	Assunto		O

	Cobertura – Geografia	Nome geográfico do local	O'
	Cobertura – Período de tempo		O'
	Cobertura – data: data de início / data de fim		O'
	Tipo de recurso	Natureza ou género do conteúdo do recurso	O
	Recurso baseado em / cópia de		O'
	Recurso é parte de		O'
	Recurso faz referência a		O'
	Recurso é referenciado por		O'
	Recurso requer		O'
	Ligação à declaração de direitos de autor		O'
<b>Documento</b>	Endereço <i>Web</i> ou URL		O
	Data de criação do recurso		O'
	Data da última modificação do recurso		O'
	Editor	Entidades envolvidas na publicação ou distribuição do recurso. Ver entidade editor	
Responsabilidade	Nome		O
	Função		O
	Afiliação		O'
	Endereço electrónico / URL pessoal		O
Editor: entidades envolvidas na publicação ou distribuição do recurso	Nome		O
	Função		O
	Afiliação		O'
	Endereço electrónico / URL pessoal		O

Fonte: Describing and cataloguing resources

**Tabela 22- Nível de consecução dos objectivos universais dos sistemas de informação bibliográfica em HUMBUL**

<b>Objectivos</b>	<b>Indicadores</b>	<b>HUMBUL</b> Legenda: S = Verifica na totalidade N= Não Verifica na totalidade SN= Verifica não totalmente
Encontrar uma entidade singular (capacidade de discriminação entre recursos semelhantes)	Atributo Complemento de título	S
	Atributo Título alternativo	S
	Atributo Versão	N
	Atributo autor (atributos que permitem distinguir autores semelhantes)	S
Identificar e seleccionar uma entidade apropriada às necessidades do utilizador	Atributo Língua	S
	Atributo Formato	N
	Atributo Requisitos de Acesso	N
	Atributo Descrição	S

	Atributo Assuntos	S
	Atributo Público-alvo	S
	Atributo Tipologia do recurso	S
	Atributo Data	S
	Atributo Versão	N
	Atributo Cobertura	S
	Atributo Editor	S
Localizar todos os documentos da mesma <b>Obra</b>	Atributo Título Uniforme	N
Localizar todos os documentos da mesma versão	Ocorrência dos tipos de relação: "É versão de" e "Tem versão"	N
Localizar todos os recursos de um mesmo autor	Controlo de autoridade autor ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas dos nomes)	N
Localizar todos os recursos com o mesmo assunto	Vocabulário normalizado e controlado	S
	Controlo de autoridade assunto ou de outra qualquer forma de controlo (ligações transparentes entre formas alternativas de designação de um mesmo conceito)	N
Obter acesso à entidade descrita	Atributo URL / URN / URI	S
	Atributo ISBN / ISSN	N
Navegabilidade no universo bibliográfico	Atributo Relação	SN
	Ocorrência de estruturas organizativas que propiciem <i>browsing</i> dos recursos	S

**Tabela 23 – Assuntos Principais e Subdivisões em SOSIG**

<b>Assunto Principal</b>	<b>Nº de Assuntos - Sub-Nível 1</b>	<b>Nº de Assuntos - Sub-Nível 2</b>	<b>Nº de Assuntos - Sub-nível 3</b>	<b>Nº de Assuntos - Sub-Nível 4</b>	<b>Total de Assuntos</b>
Business	5	19	14	---	38
Economics	13	40	-----	---	53
Education	12	8	--	---	20
Environmental Sciences	7	8	--	---	15
Ethnology, Ethnography, Anthropology	13	33	--	---	46
European Studies	3	6	47	----	56
Geography	9	2	--	---	11
Government and Public Administration	16	39	2	---	57
Law	3	287	117	4	411
Philosophy	10	6	--	--	16
Politics	13	42	10	--	65
Psychology	18	109	--	--	127
Social Science General	2	2	--	--	4
Social Welfare	19	1	----	----	20
Sociology	24	2	---	---	26

Statistics	7	---	----	---	7
Women's Studies	15	---	---	---	15
Totais	189	604	190	4	987

**Tabela 24 - Distribuição dos recursos classificados em cada um dos níveis da estrutura em SOSIG.**

<b>Temas Principais</b>	<b>Nível principal</b>	<b>1º nível</b>	<b>2º nível</b>	<b>3º nível</b>	<b>4º nível</b>	<b>Total</b>
Business	345	949	1.515	821	-	3.630
Economics	282	1.667	4.817			6.766
Education	322	902	226			1.450
Environmental sciences and issues	104	1.092	396			1.592
Ethnology, Ethnography, Anthropology	95	886	1.439			2.420
European Studies	0	1.239	577	577	-	2.393
Geography	112	582	98	-	-	792
Government and Public Administration	392	2.373	3.018	140		5.923
Law	83	479	5.165	730	31	6.488
Philosophy	121	472	237			830
Politics	217	1.919	4.354	627	-	7.117
Psychology	76	499	1.373	-	-	1.948
Social Science General	407	563	147	-	-	1.117
Social Welfare	108	2.627	109	-	-	2.844
Sociology	169	3.229	253	-	-	3.651
Statistics	311	809		-	-	1.120
Women Studies	296	835		-	-	1.131
Totais	3.440	21.122	23.724	2.895	31	51.212

**Tabela 25- Distribuição dos recursos pelos assuntos na escala em SOSIG**

<b>Nº de recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1-10	367	38.7
11-20	144	15.2
21-30	60	6.3
31-40	39	4.1
41-50	41	4.3
51-60	40	4.2
61-70	35	3.7
71-80	25	2.6
81-90	23	2.4
91-100	19	2
101-150	63	6.6
151-200	40	4.2

+200	53	5.5
------	----	-----

**Tabela 26 - Distribuição dos recursos pelos assuntos no intervalo 1-5 recursos em SOSIG**

Nº de recursos	Nº de Assuntos Frequência absoluta	Nº de Assuntos Frequência relativa
1	81	30.5
2	67	25.3
3	46	17.3
4	35	13.2
5	36	13.6
Total	265	

**Tabela 27 - Distribuição dos recursos pelos temas (sem atribuição múltipla) em HUMBUL**

Tema	Nº de Recursos Frequência absoluta	Nº de Recursos Frequência relativa %
<i>Archaeology</i>	694	14.2
<i>History</i>	1178	24.2
<i>History &amp; Philosophy of Science</i>	217	4.4
<i>American Studies</i>	278	5.7
<i>Linguistics</i>	202	4.1
<i>German Studies</i>	79	1.6
<i>Spanish Studies</i>	39	0.8
<i>Russian and Slavonic Studies</i>	65	1.3
<i>Classics</i>	259	5.3
<i>Philosophy</i>	497	10.2
<i>Religion &amp; Theology</i>	510	10.4
<i>Humanities computing</i>	144	2.9
<i>English Studies</i>	577	11.9
<i>French Studies</i>	58	1.2
<i>Italian Studies</i>	37	0.8
<i>Comparative Literature</i>	34	0.7
Total	4868	

**Tabela 28 – Assuntos Principais e Subdivisões em PSIGATE**

Assunto Principal	Nº de Assuntos- Sub-Nível 1	Nº de Assuntos - Sub-Nível 2	Total de assuntos
<i>Astronomy</i>	9	31	40
<i>Chemistry</i>	12	42	54
<i>Earth Sciences</i>	19	30	49
<i>Sciences History / Policy</i>	12	---	12
<i>Materials Sciencies</i>	10	3	13
<i>Physics</i>	18	39	57
Total	80	145	225

**Tabela 29 – Comparação do nível de especificidade da representação dos assuntos na estrutura de assuntos e na Classificação Decimal de Dewey em PSIGATE**

**Assunto : Physics**

<b>1º Nível</b>	<b>CDD</b>	<b>2º Nível</b>	<b>CDD</b>
<i>Acoustics</i>	534		
<i>Condensed matter physics</i>	530.41		
<i>Crystallography</i>	548		
<i>Electricity and magnetism</i>	537	<i>Magnetic resonance</i>	538.36
		<i>Magnetism</i>	538
		<i>Optoelectronic</i>	537.5
		<i>Semiconductors</i>	537.622
		<i>Superconductors</i>	537.623
		<i>Electromagnetism</i>	537.1
<i>Fluid Mechanics</i>	532		
<i>General Physics</i>	530		
<i>Heat and thermodynamics</i>	536		
<i>Mathematical physics</i>	530.15		
<i>Measurement and instruments</i>	530.8	<i>Chromatography Instruments</i>	530.82
		<i>Meters</i>	530.82
		<i>Microscopes</i>	530.81
		<i>Other instruments</i>	530.81
		<i>Particle accelerators and detectors</i>	530.81
		<i>Photography</i>	530.81
		<i>Sensors</i>	530.81
		<i>Spectrometry Instrumentation</i>	530.82
		<i>Spectroscopy Instrumentations</i>	530.83
		<i>Telescopes</i>	530.81
		<i>Units</i>	530.8
		<i>Mechanics, statics and dynamics</i>	531
<i>Molecular, atomic and particle physics</i>	539	<i>Antimatter</i>	539.721
		<i>Atomic structure</i>	539.14
		<i>CP violation</i>	539.721
		<i>Cosmic Rays</i>	539.722
		<i>Dark Matter</i>	523.112
		<i>Hadrons</i>	539.721
		<i>Neutrinos</i>	539.721
		<i>Neutron scattering</i>	539.721
		<i>Particle acceleration and detection</i>	539.73
<i>Protons</i>	539.721		
<i>Nuclear physics</i>	539.75		
<i>Optics and Light</i>	535	<i>Colour</i>	535.6
		<i>Fibre optics</i>	535.89
		<i>Lasers</i>	535.59
		<i>Nonlinear Optics</i>	535.2
		<i>Quantum Optics</i>	535.14
		<i>Solitons</i>	535.2
		<i>X- Rays</i>	535.8
<i>Physics of activities and environments</i>	530	<i>Physics of food and drink</i>	664
		<i>Sports Physics</i>	530
<i>Solids, liquids, gases and plasmas</i>	530.4		

<i>Theoretical physics</i>	530.1	<i>Quantum Physics</i>	530.12
		<i>Relativity</i>	530.11
<i>Time</i>	529		
<i>General science</i>	500	<i>Physical constants and units</i>	500

**Tabela 30 - Distribuição dos recursos classificados em cada um dos níveis da estrutura de *browsing* em PSIGATE**

<b>Tema</b>	<b>1º nível</b>	<b>2º Nível</b>	<b>Total</b>
<i>Astronomy</i>	556	469	1025
<i>Chemistry</i>	1263	1430	2693
<i>Earth Sciences</i>	1282	432	1714
<i>Materials Science</i>	794	35	829
<i>Physics</i>	1798	862	2660
<i>Science History / Policy</i>	674	--	674
<i>General Science</i>	150	47	197
TOTAIS	6517	3275	9792

**Tabela 31 - Recursos classificados em cada nível de acordo com a escala em PSIGATE**

<b>Nº de Recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1-10	63	27.6
11-20	40	17.5
21-30	18	7.9
31-40	28	12.3
41-50	8	3.5
51-60	8	3.5
61-70	13	5.7
71-80	8	3.5
81-90	8	3.5
91-100	9	3.9
101-150	17	7.5
151-200	5	2.2
+200	3	1.3

**Tabela 32 - Distribuição dos recursos pelos assuntos no intervalo 1-5 em PSIGATE**

<b>Nº de Recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1	5	17.2
2	8	27.6
3	5	17.2
4	4	13.8
5	7	24.1
Total	29	100

**Tabela 33 – Assuntos Principais e Subdivisões no Directório AgriFor**

<b>Assunto Principal</b>	<b>Nº de Assuntos Sub-Nível 1</b>	<b>Nº de Assuntos Sub-Nível 2</b>	<b>Nº de Assuntos-Sub-Nível 3</b>	<b>Total de Assuntos</b>
<i>Agriculture</i>	8	35	16	59
<i>Livestock</i>	10	45	35	90
<i>Plant diseases and Pest Control</i>	9	7	---	16
<i>Fisheries</i>	6	15	---	21
<i>Forestry</i>	8	51	47	106
<i>Crops and Horticulture</i>	5	21	19	45
<i>Soil Science</i>	11	6	---	17
<i>Natural Resources and Plant Science</i>	16	16	4	36
<i>Economics, Trade &amp; Rural Development</i>	10	22	15	47
<i>Food</i>	8	40	8	56
<b>Total</b>	<b>91</b>	<b>258</b>	<b>144</b>	<b>493</b>

**Tabela 34 - Distribuição dos Recursos atribuídos em cada um dos níveis em Agrifor**

<b>Assunto Principal</b>	<b>1º Nível</b>	<b>2º Nível</b>	<b>3º Nível</b>	<b>Total</b>
<i>Agriculture</i>	17	528	241	786
<i>Livestock</i>	0	205	238	443
<i>Plant diseases and Pest control</i>	74	78	--	152
<i>Fisheries</i>	3	117	--	120
<i>Forestry</i>	0	232	461	693
<i>Crops and Horticulture</i>	1	166	99	266
<i>Soil Science</i>	83	14	--	97
<i>Natural resources and Plant Science</i>	95	188	211	480
<i>Economic &amp; Trade and Rural development</i>	81	188	211	480
<i>Food</i>	39	300	137	476
<b>Totais</b>	<b>393</b>	<b>1927</b>	<b>1598</b>	<b>3783</b>

**Tabela 35 - Distribuição dos recursos pelos assuntos de acordo com escala em Agrifor**

<b>Nº de Recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1-10	269	72.3
11-20	60	16.1
21-30	15	4
31-40	13	3.5
41-50	5	1.3
51-60	4	1
61-70	2	0.5
71-80	1	0.3

81-90	2	0.5
91-100	0	---
101-150	1	0.3
151-200	0	---
+200	0	---

**Tabela 36 - Distribuição do nº de recursos nos assuntos no intervalo 1-5 em Agrifor**

<b>Nº de Recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1	59	28.8
2	53	25.9
3	32	15.6
4	34	16.6
5	27	13.1
Total	205	100

**Tabela 37 - Distribuição do nº de recursos nos assuntos no intervalo 1-5 em Agrifor**

<b>Nº de Recursos</b>	<b>Nº de Assuntos Frequência absoluta</b>	<b>Nº de Assuntos Frequência relativa</b>
1	59	28.8
2	53	25.9
3	32	15.6
4	34	16.6
5	27	13.1
Total	205	100

**Tabela 38 - Distribuição dos recursos por cada assunto em Agrifor**

<b>Assunto</b>	<b>Nº de recursos classificados Frequência absoluta</b>	<b>Nº de recursos classificados Frequência relativa</b>
<i>Agriculture</i>	786	20.7
<i>Livestock</i>	443	11.7
<i>Plant diseases and Pest control</i>	152	4
<i>Fisheries</i>	120	3.2
<i>Forestry</i>	693	18.3
<i>Crops and Horticulture</i>	266	7
<i>Soil Science</i>	97	2.6
<i>Natural resources and Plant Science</i>	480	7.1
<i>Economic &amp; Trade and Rural development</i>	480	12.7
<i>Food</i>	476	12.6

**Tabela 39 - Exemplo da Lista de Assuntos baseada na CDD com a respectiva notação em Agrifor**

<b>Assunto Principal</b>	<b>CDD</b>
<i>Administration of economy &amp; environment</i>	354
<i>Agriculture &amp; related technologies</i>	630
<i>Animal husbandry</i>	636
<i>Applied physics</i>	621
<i>Beverage technologies</i>	663
<i>Bibliographies</i>	011
<i>Biochemistry</i>	572
<i>Building materials</i>	691
<i>Chemical engineering</i>	660

**Tabela 40 – Notações utilizadas por classe da CDD em Agrifor**

<b>Classe CDD</b>	<b>Nº de Notações Utilizadas Frequência absoluta</b>	<b>Nº de Notações utilizadas Frequência relativa</b>
0	2	3.2
3	20	31.7
5	12	19
6	27	42.8
7	2	3.2

**Tabela 41 - Comparação entre as termos utilizados nas diversas formas de *browsing* em Agrifor  
Assunto escolhido: DOENÇAS – Termo: Diseases**

<b>CDD</b>	<b>Directório – 1º nível</b>	<b>Directório – 2º nível</b>	<b>CABI – 1º nível</b>	<b>CABI – 2º Nível</b>	<b>CABI – 3º nível</b>
<i>Diseases</i>	<i>Diseases</i>	<i>General</i>	<i>Diseases</i>	<i>Animal diseases</i>	<i>Bee diseases</i>
					<i>Cattle diseases</i>
					<i>Fish diseases</i>
					<i>Poultry diseases</i>
					<i>Sheep diseases</i>
					<i>Swine diseases</i>
		<i>Bacterial diseases</i>		<i>Foul brod</i>	
				<i>Salmonellosis</i>	
		<i>Fungus diseases</i>			
		<i>Viral diseases</i>			
		<i>Plant diseases</i>		<i>Forest decline</i>	
				<i>Fungal diseases</i>	
<i>Root rots</i>					
<i>Wilts</i>					
<i>Postharvest decay</i>					
<i>Prior diseases</i>	<i>Creutzfeld-Jakob disease</i>				
	<i>Bovine spongiform encephalopathy</i>				
	<i>Scrapie</i>				

**Tabela 42 – Assuntos Principais e subdivisões em Natural**

<b>Assunto Principal</b>	<b>CDD</b>	<b>Sub-Nível 1</b>	<b>CDD</b>	<b>Sub-Nível 2</b>	<b>CDD</b>	<b>Sub-Nível 3</b>	<b>CDD</b>	<b>Sub-nível 4</b>	<b>Total</b>			
<i>Natural resources &amp; energy</i>	333.7	<i>Biological resources - biodiversity</i>	333.95	---	---	---	---	---	2			
		<i>Conservation &amp; Protection</i>	333.72									
<i>Paleontology</i>	560	<i>Fossil fishes</i>	Não utilizado	<i>Reptils (dinosaurs)</i>	567.9	---	---	---	4			
		<i>Fossil Plants &amp; microorganisms</i>	561									
		<i>Paleontological organizations</i>	560.6									
<i>Biology</i>	570	<i>Biological organisations</i>	570.6	-----	---	---	---	18				
		<i>Biologists</i>	570.92	----	----	---	---					
		<i>Ecology</i>	<i>Ecological organisations</i>	577	577.06	---	---		---			
			<i>Forest ecology</i>							577.3	---	---
			<i>Freshwater ecology</i>							577.6	---	---
			<i>Marine ecology</i>							577.7	---	---
		<i>Genetic &amp; evolution</i>	Não usado	<i>Evolution</i>	576.8	---	---					
		<i>Microorganisms, fungi, algae</i>	579	<i>Fungi</i>	579.5	---	---					
		<i>Natural history of organisms</i>	578			<i>Geographic treatment</i>	Não usado		<i>Arctic Islands &amp; Antarctica</i>	578.0998		
									<i>North America</i>	578.097		
									<i>Natural historians, naturalists</i>	578.092	----	---
<i>Natural history organisations</i>	578.06	----	---									

		<i>Physiology</i>	571	---	---	---	---		
<i>Botany- Plants</i>	580	<i>Botanical organizations</i>	580.6	<i>Living collections</i>	580.73	<i>Australasia</i>	580.739	2	24
						<i>Europe</i>	580.734		
						<i>North America</i>	580.737		
				<i>Specimen collections</i>	Não usado	<i>Australasia</i>	580.749		
						<i>Europe</i>	580.744		
						<i>North America</i>	580.747		
		<i>Geographic treatment - Regional flora</i>	Não usado	<i>Africa</i>	581.96	---	---		
				<i>Asia</i>	581.95	---	---		
				<i>Australasia</i>	581.99	---	---		
				<i>Central America</i>	581.972	---	---		
				<i>Europe</i>	581.94	---	---		
				<i>North America</i>	581.97	---	---		
		<i>Habitat groups</i>	Não usado	<i>Alpine Plants</i>	581.7538	---	---		
<i>Aquatic Plants</i>	581.76			---	---				
<i>Plant genetics &amp; evolution</i>	581.3	---							
<i>Rare &amp; endangered plants</i>	581.68	---							
<i>Zoology animals</i>	- 590	<i>Geographic treatment</i>	Não usado	<i>Australiasa</i>	591.99	---		1	15
				<i>Europe</i>	591.94	---			
				<i>North America</i>	591.97	---			
		<i>Rare &amp; endangered animals</i>	591.68	---	---	---			
		<i>Zoological organisations</i>	590.6	<i>Living collections inc. in Zoos</i>	Não usado	<i>Europe</i>	Não usado		
						<i>Specimen collections</i>	Não usado	<i>Asia</i>	590.745
				<i>Europe</i>	590.744				
				<i>North America</i>	590.747				
		<i>South America</i>	590.748						
<i>Zoologists</i>	590.92	---	---						
<b>Total</b>		21		26		13		3	63

**Tabela 43 – Distribuição dos recursos classificados em cada um dos níveis da estrutura em Natural**

<b>Tema</b>	<b>1º nível</b>	<b>2º nível</b>	<b>3º nível</b>	<b>4º nível</b>	<b>Total</b>
<i>Botany – Plants</i>	1108	580	50	--	1738
<i>Ecology</i>	293	173	20	--	486
<i>Mycology</i>	112	29	3	--	144
<i>Protistology</i>	85	9	1	--	95
<i>Zoology – Animals</i>	1341	586	56	--	1983
<i>Biology</i>	91	--	--	--	91
<i>Evolution</i>	115	14	--	--	129
<i>Palaeobotany</i>	168	64	4	--	236
<i>Palaeozoology</i>	315	140	14	--	469
<i>Stratigraphic Palaeontology</i>	129	58	--	--	187
<i>Biodiversity</i>	124	64	12	--	200
<i>Biogeography</i>	34	4	--	--	38
<i>Conservation &amp; Protection</i>	576	285	29	--	890
<i>Environnemental Sciences</i>	136	--	--	--	136
<i>Systematics &amp; Taxonomy</i>	264	60	--	--	324
<i>Oceanography &amp; Aquatic Sciences</i>	353	350	64	5	772
Totais	5.244	2.416	253	5	7918

**Tabela 44 – Distribuição dos recursos pelos grandes temas em Natural**

<b>Tema</b>	<b>Nº de recursos classificados (distribuição absoluta)</b>	<b>Nº de recursos classificados (distribuição relativa)</b>
<i>Botany – Plants</i>	1738	21.9
<i>Ecology</i>	486	6.1
<i>Mycology</i>	144	2
<i>Protistology</i>	95	1.2
<i>Zoology – Animals</i>	1983	25
<i>Biology</i>	91	1.1
<i>Evolution</i>	129	1.6
<i>Palaeobotany</i>	236	3
<i>Palaeozoology</i>	469	5.9
<i>Stratigraphic Palaeontology</i>	187	2.4
<i>Biodiversity</i>	200	2.5
<i>Biogeography</i>	38	0.5
<i>Conservation &amp; Protection</i>	890	11.2
<i>Environnemental Sciences</i>	136	1.7
<i>Systematics &amp; Taxonomy</i>	324	4
<i>Oceanography &amp; Aquatic Sciences</i>	772	9.7
Totais	7918	100

**Tabela 45 - Descritores e respectivas notações em BIORES**

<b>Descritores</b>	<b>Notação LC</b>	<b>Notação NLM</b>
<i>Anatomy</i>	QL800	
<i>Bacteria</i>	QR74.8	
<i>Bibliographic databases</i>	Z699	
<i>Biochemistry</i>	QP500	
<i>Biological Sciences</i>	QH301	
<i>Biophysics</i>	QH 505	
<i>Biotechnology</i>	TP248	
<i>Botany</i>	QK1	
<i>Cell Biology</i>	QH 573	
<i>Ecology</i>	QH540	
<i>Education, schools and departments</i>		W19
<i>Embryology</i>	QL951	
<i>Evolution</i>	QH355	
<i>Fungi</i>	QK600	
<i>General Science</i>	Q1	
<i>Genetics</i>	QH426	
<i>Histology</i>		QS504
<i>Immunology</i>	QR179	
<i>Journals and other publications</i>		W1
<i>Laboratory Animals</i>		QY50
<i>Microbiology</i>	QR1	
<i>Microscopy</i>	QH200	
<i>Molecular Biology</i>	QH506	
<i>Neoplasm's</i>		QZ200
<i>Nuerosciences</i>	QP350	
<i>Parasitology</i>		QX4
<i>Physiology</i>	QP1	
<i>Software and Tools</i>	QA76.75	
<i>Toxicology</i>		QV600
<i>Virology</i>	QR355	
<i>Zoology</i>	QL1	

**Tabela 46 - Distribuição dos recursos pelos assuntos de acordo com escala em Vetgate**

<b>Nº de Recursos</b>	<b>Nº de Assuntos. Frequência absoluta</b>	<b>Nº de Assuntos. Frequência relativa (%)</b>
1-10	53	58.2
11-20	10	11
21-30	7	7.7
31-40	5	5.5
41-50	3	3.3
51-60	4	4.4
61-70	--	--
71-80	--	--
81-90	--	--
91-100	1	1
101-150	7	7.7
151-200	1	1
+200	--	--

Total	91	100
-------	----	-----

**Tabela 47 – Distribuição dos recursos pelos assuntos em Vetgate**

Nº de recursos classificados	Nº de assuntos Frequência absoluta	Nº de assuntos Frequência relativa (%)
1	14	34
2	9	22
3	6	14.6
4	6	14.6
5	6	14.6
Total	41	

**Tabela 48 - Distribuição dos recursos pelos assuntos de acordo com escala em OMNI**

Nº de Recursos	Nº de Assuntos. Frequência absoluta	Nº de Assuntos. Frequência relativa (%)
1-10	15	17.8
11-20	4	4.8
21-30	5	6
31-40	4	4.8
41-50	4	4.8
51-60	3	3.6
61-70	4	4.8
71-80	7	8.3
81-90	2	2.4
91-100	--	--
101-150	10	12
151-200	7	8.3
+200	19	22.6
Total	84	100

**Tabela 49 - Exemplo da utilização de ACM computing classification para a classe principal "Computing methodologies" (3 primeiras sub-classes) em EEVL**

EEVL - Assunto principal	Notações ACM	EEVL – 1º nível de subdivisão	Notações ACM	EEVL – 2º nível de subdivisão	Notações ACM
Computing methodologies	I	Computing methodologies (General)	I.0		
		<i>Symbolic and algebraic manipulation</i>	I.1	<i>Symbolic and algebraic manipulation (General)</i>	I.1.0
				<i>Algorithms</i>	I.1.1

				<i>Language Systems</i>	I.1.3
		<i>Artificial intelligence</i>	I.2	<i>Artificial intelligence (General)</i>	I.2.0
				<i>Application and expert systems</i>	I.2.1
				<i>Automatic programming</i>	I.2.2
				<i>Deduction and Theorem proving</i>	I.2.3
				<i>Knowledge representation, formalisms and methods</i>	I.2.4
				<i>Programming languages and software</i>	I.2.5
				<i>Learning</i>	I.2.6
				<i>Natural language processing</i>	I.2.7
				<i>Problem solving, control methods and search</i>	I.2.8
				<i>Robotics</i>	I.2.9
				<i>Vision and scene understanding</i>	I.2.10
		<i>Distributed artificial intelligence</i>	I.2.11		
		<i>Computer graphics</i>	I.3	<i>Computer graphics (General)</i>	I.3.0
				<i>Hardware architecture</i>	I.3.1
				<b>(Inexistente)</b>	I.3.2
				<b>(Inexistente)</b>	I.3.3
				<i>Graphics utilities</i>	I.3.4
				<i>Computational geometry and object modeling</i>	I.3.5
		<i>Methodology and techniques</i>	I.3.6		

				<i>Three-dimensional graphics and realism</i>	I.3.7.
				<i>Applications</i>	I.3.8.

**Tabela 50 – Níveis de Subdivisão dos Assuntos em EEVL**

<b>Assunto Principal</b>	<b>Nº de Assuntos - Sub-nível 1</b>	<b>Nº de Assuntos - Sub-Nível 2</b>	<b>Nº de Assuntos - Sub-Nível 3</b>	<b>Nº de Assuntos - Sub-nível 4</b>	<b>Total de Assuntos</b>
<i>Engineering</i>	15	69	103	57	244
<i>Mathematics</i>	10	53	96	---	159
<i>Computing</i>	11	67	163	--	241
<b>Total</b>	36	189	362	57	644

**Tabela 51 – Distribuição dos recursos classificados em cada um dos níveis da estrutura de *browsing* em EEVL**

<b>Tema</b>	<b>1º Nível</b>	<b>2º Nível</b>	<b>3º Nível</b>	<b>4º Nível</b>	<b>Total</b>
<i>Engineering</i>	945	7821	4069	2226	15061
<i>Mathematics</i>	0	1874	1002	--	2876
<i>Computing</i>	129	831	1896	--	2856
Totais	1074	10526	6967	2226	20793

**Tabela 52 – Número de Recursos classificados em cada assunto em EEVL**

<b>Nº de Recursos</b>	<b>Nº de Assuntos. Frequência absoluta</b>	<b>Nº de Assuntos. Frequência relativa (%)</b>
1-10	268	50
11-20	54	10
21-30	44	8.2
31-40	26	4.8
41-50	21	3.9
51-60	16	3
61-70	12	2.2
71-80	15	2.8
81-90	5	0.9
91-100	9	1.7
101-150	32	6
151-200	15	2.8
+200	19	3.5
Total	536	

**Tabela 53 – Distribuição dos recursos pelos assuntos em EEVL**

Nº de recursos	Nº de Assuntos Frequência absoluta	Nº de Assuntos Frequência relativa (%)
1	68	38
2	35	19.5
3	34	19
4	33	18.4
5	9	5
Total	179	

**Tabela 54 - Análise de conteúdo do elemento descrição nas Regras de Descrição dos Serviços**

Categorias	HUMBUL	SOSIG	PSIGATE BIOME	EEVL
Tipologia do recurso	S	S	S	S
Responsabilidade	S	S	S	S
Cobertura temática do recurso	S	S	S	S
Limites temporais e geográficos da informação	S	S	N	N
Informação em outras línguas para além da original	S	S	S	S
Formatos	N	N	S	S
<i>Mirror sites</i>	N	N	S	N
Condicionantes tecnológicas e económicas no acesso à informação	N	S	S	S
Ligação a outros recursos	N. Remete para o elemento ligação	N	S	S
Extensão	Média: 100 Máximo: 300 palavras	N	100/150 palavras	N
Estrutura	Texto não estruturado. Evitar listas enumerativas	1 parágrafo.		N

Fonte: Cataloguing Guidelines

Legenda: S= Verifica N= Não Verifica

**Tabela 55 – Extensão dos Resumos**

Serviço	Média (valores absolutos)
HUMBUL	168.6
EEVL	100.7
SOSIG	90.6
BIOME	79.1

PSIGATE	71.7
---------	------

Fonte: amostra seleccionada

**Tabela 56 – Complexidade das Palavras (número médio dos caracteres por palavra)**

<b>Serviço</b>	<b>Média</b>
PSIGATE	5.7
EEVL	5.7
SOSIG	5.5
BIOME	5.4
HUMBUL	5.3

**Tabela 57 – Complexidade das frases (número de palavras por frase)**

<b>Serviço</b>	<b>Média</b>
HUMBUL	24.6
PSIGATE	24
BIOME	22.3
EEVL	22
SOSIG	21

**Tabela 58 – Frases na Voz Passiva (%)**

<b>Serviço</b>	<b>Média</b>
BIOME	29.8%
HUMBUL	21%
PSIGATE	20.6%
EEVL	18.4%
SOSIG	17.8%

**Tabela 59 – Flesh Reading Ease**

<b>Serviço</b>	<b>Média</b>
HUMBUL	26.7
BIOME	21.7
SOSIG	21.5
EEVL	20.1
PSIGATE	17.8

**Tabela 60 – Relação entre os resultados do teste FRE e a extensão dos resumos**

<b>FRE</b>	<b>Extensão do Resumo</b>
HUMBUL	168.6
BIOME	79.1
SOSIG	90.6

EEVL	100.7
PSIGATE	71.7

**Tabela 61 – Relação entre os resultados do teste FRE e a extensão da frase**

<b>FRE</b>	<b>Extensão da Frase</b>
HUMBUL	24.6
BIOME	22.3
SOSIG	21
EEVL	22
PSIGATE	24

**Tabela 62 – Relação entre os resultados do teste FRE e a complexidade da frase**

<b>FRE</b>	<b>Complexidade da Frase (em de caracteres por palavra)</b>
HUMBUL	5.3
BIOME	5.4
SOSIG	5.5
EEVL	5.7
PSIGATE	5.7