

**Organização do conhecimento e representação de assuntos:
caminhos para uma efectiva recuperação da informação em
ambientes de rede**

Dissertação de Mestrado em Ciências Documentais, apresentada à Universidade de Évora.

Esta dissertação não inclui as críticas e sugestões feitas pelo júri

Autora: Maria Luísa Ferreira Nunes dos Santos

Orientadora: Fernanda Maria Guedes de Campos

UNIVERSIDADE DE ÉVORA

2005

**Organização do conhecimento e representação de assuntos:
caminhos para uma efectiva recuperação da informação em
ambientes de rede**

*Dissertação de Mestrado em Ciências Documentais, apresentada à Universidade de Évora.
Esta dissertação não inclui as críticas e sugestões feitas pelo júri*



155006

Autora: Maria Luísa Ferreira Nunes dos Santos
Orientadora: Fernanda Maria Guedes de Campos

UNIVERSIDADE DE ÉVORA

2005

Organização do conhecimento e representação de assuntos: caminhos para uma efectiva recuperação da informação em ambientes de rede

Resumo

O impacto das TIC, da Internet e da WWW obriga as bibliotecas a repensar os seus métodos, processos e circuitos, a criar novos modelos de desenvolvimento de serviços e a implementar mudanças, por vezes profundas, a nível interno e inter-organizacional. Projectos como o RENARDUS, o MACS e o HILT, que desenvolveram experiências tendentes à melhoria da recuperação a um nível global (na Web, num ambiente multilíngue e simultaneamente em bibliotecas, arquivos e museus), podem apontar caminhos para uma efectiva recuperação por assunto em ambientes de rede, por mediação dos catálogos de bibliotecas, arquivos e museus ou através da Web. Da análise dessas experiências procedeu-se à sua adaptação a catálogos colectivos como a PORBASE – Base Nacional de Dados Bibliográficos. Procura-se, pela análise das diversas questões inerentes à normalização e à interoperabilidade e pela redefinição de estratégias de cooperação, estabelecer caminhos com o objectivo de melhorar o acesso à informação, nomeadamente aos seus conteúdos temáticos.

Knowledge organisation and subject representation: ways for effective information retrieval in the web environment

Abstract

Technology, Internet and the WWW forced libraries to rethink their methods and processes, to create new models for service development and to implement (sometimes deep) changes, both internal and inter-organisational. Union catalogues like PORBASE (Portuguese National Bibliographic Database) must adapt themselves to this new reality, by analysing standardisation and interoperability issues and redefining cooperative strategies to improve information access, namely to subject contents. Subject retrieval possibilities through library, archives and museums catalogues and through the Web have been studied by projects like RENARDUS, MACS and HILT. These projects have been developing some experiences in order to improve global retrieval (in the Web, in a multilingual environment and simultaneously in libraries, archives and museums) and the methods they followed can be useful in finding new ways for subject retrieval in union catalogues like PORBASE.

Agradecimentos

Quero em primeiro lugar agradecer à minha orientadora, Dr.^a Fernanda Maria Guedes de Campos, pela paciência e disponibilidade sempre demonstradas, pelo cuidado com que leu estas páginas e pelas sugestões de melhoria que fez ao texto final. Agradeço também à minha colega e amiga Prof. Doutora Ana Cardoso de Matos, sempre disponível e cujo apoio foi muito importante para a concretização deste trabalho. Um agradecimento especial vai também para a Biblioteca Nacional, pelas facilidades concedidas e para as pessoas que comigo trabalham na Divisão da PORBASE por terem assegurado, nas minhas curtas ausências, os trabalhos em curso. Agradeço ainda à Comissão de Curso do Mestrado, particularmente à Prof. Doutora Fernanda Olival, pelo interesse com que acompanhou todo o processo. Finalmente, um agradecimento aos meus amigos mais próximos e à minha família, em particular aos meus pais, pela compreensão e apoio sempre presentes.

Sumário

1. Introdução	1
2. Estrutura da informação em bases de dados cooperativas, ou catálogos colectivos	8
2.1 A norma ISO 2709 e os formatos MARC na génese das bases de dados bibliográficos.....	8
2.2 Os sistemas de suporte às bases de dados bibliográficos	9
2.3 Cooperação e normalização em bases de dados: o registo bibliográfico e o registo de autoridade.....	15
2.4 Características gerais das bases de dados cooperativas ou catálogos colectivos	19
2.5 Os catálogos colectivos e os seus modelos	23
3. A PORBASE – Base Nacional de Dados Bibliográficos: breve caracterização de um catálogo colectivo	29
3.1 A criação da PORBASE e algumas etapas do seu desenvolvimento como projecto cooperativo.....	29
3.2 Caracterização do conteúdo da base de dados e do universo das bibliotecas cooperantes	34
3.3 A PORBASE e os modelos de catálogo colectivo	38
4. O acesso por assunto à informação	40
4.1 Os modelos tradicionais de representação do conhecimento e de recuperação de assuntos	40
4.1.1 As classificações universais	40
4.1.2 Os sistemas terminológicos: tesouros e listas de cabeçalhos de assunto	43
4.1.3 Outros sistemas.....	46
4.2 O acesso por assunto nos catálogos colectivos ou bases de dados cooperativas	49
4.3 A recuperação por assuntos na <i>Web</i>	52
4.3.1 Os portais temáticos (" <i>Subject gateways</i> ").....	54
4.3.2 As ontologias	57
4.3.3 Outras iniciativas para a organização do conhecimento na <i>Web</i>	63
5. Três abordagens complementares à questão do acesso por assunto.....	65
5.1 O Projecto RENARDUS – um exemplo de organização do conhecimento e de interoperabilidade entre diferentes sistemas de classificação e um modelo comum.....	65
5.2 MACS (Multilingual Access to Subjects) – um exemplo de interoperabilidade entre linguagens de indexação de diferentes bibliotecas nacionais, em diferentes línguas.....	76
5.3 O projecto HILT (High Level Thesaurus) – uma proposta de interoperabilidade entre diferentes linguagens de indexação utilizadas por entidades tipologicamente distintas (bibliotecas, arquivos, museus).....	81
6. Uma proposta de desenvolvimento da PORBASE no contexto do acesso por assuntos à informação em rede	90
7. Conclusão	102
Bibliografia	107
Anexos	130
Anexo 1 – Registos de autoridade (Unimarc/Autoridades).....	131
Anexo 2 – Quadro genealógico das linguagens documentais.....	132
Anexo 3 – Construir uma ontologia.....	133
Anexo 4 – Elementos e qualificadores de metadados que fazem parte do perfil de aplicação Renardus (Maio 2001)	135
Anexo 5 – Arquitectura do projecto-piloto RENARDUS.....	137
Anexo 6 – Ferramenta de mapeamento Renardus.....	138
Anexo 7 – Estrutura gráfica de navegação Renardus.....	139
Anexo 8 – Modelo organizacional para o serviço intermediário RENARDUS.....	140
Anexo 9 – Projecto MACS – modelos de ligação.....	142
Anexo 10 – Arquitectura do Projecto MACS.....	145
Anexo 11 – Projecto MACS – exemplos de equivalência estabelecida entre as linguagens LCSH, SWD e RAMEAU.....	146
Anexo 12 – Exemplo de pesquisa no protótipo MACS.....	147
Anexo 13 – Projecto HILT – esquemas usados pelos depositários de informação seleccionados.....	150

Anexo 14 – Projecto HILT – pontos fracos e fortes das linguagens analisadas.....	151
Anexo 15 – Projecto HILT – conjunto de opções colocadas aos participantes (depositários de informação no workshop de Junho de 2001.....)	153
Anexo 16 – Projecto HILT – diagrama TeRM (Interactive Terminologies Route Map).....	154
Anexo 17 – Siglas e acrónimos utilizados.....	155

1. Introdução

A escolha do tema desta dissertação resulta da necessidade sentida de aprofundar e clarificar algumas questões inerentes à recuperação da informação, nomeadamente a possibilidade de aceder, de uma forma fiável e consistente, aos conteúdos temáticos disponíveis em ambientes de rede caracterizados pela diversidade e heterogeneidade.

Recolher, tratar e difundir informação, contribuindo dessa forma para o desenvolvimento da investigação e do conhecimento, foi sempre a missão de organizações como as bibliotecas, os arquivos, os museus e outras entidades afins, independentemente da designação assumida ao longo do tempo. Se é verdade que os meios informáticos e as novas tecnologias vieram colocar novos desafios à comunidade alargada dos serviços de informação, é igualmente verdade que as suas funções essenciais se mantêm inalteráveis. Encontrando-nos num ponto de viragem devido ao advento e expansão das novas tecnologias da informação e da comunicação (TIC), não devemos por isso supor que tudo é agora diferente. Há, de facto, uma viragem, mas as preocupações fundamentais dos serviços de informação – selecção, processamento, difusão – mantêm-se. O impacto das TIC obriga a repensar métodos, processos, circuitos, o que sem dúvida implica a criação de novos modelos e a implementação de mudanças, por vezes profundas, a nível das organizações e das relações que estas estabelecem entre si, para que a sua missão possa concretizar-se de acordo com as novas exigências da sociedade. No caso concreto das bibliotecas, para além do aumento e da diversificação dos seus recursos informativos, é possível atingir públicos-alvo cada vez mais vastos, diversos e exigentes, já que a informação transpõe paredes e fronteiras e chega a todos os pontos do globo. À aparente simplificação dos procedimentos técnicos tradicionais nestas instituições, junta-se a complexidade induzida por esta diversidade de recursos e públicos e pela possibilidade de melhor interagir com eles, pela necessidade de desenvolver mecanismos para uma cooperação cada vez mais estreita e pela tentativa não só de responder às novas exigências, mas também de criar novas necessidades junto dos públicos/utilizadores a que se destinam os serviços,

estimulando a investigação e contribuindo assim para o avanço do conhecimento. Todas estas mudanças obrigam, por seu turno, à existência nas bibliotecas de pessoas com formação adequada à prossecução de objectivos mais ambiciosos, de modo a poderem desenvolver serviços de qualidade. Se as novas tecnologias permitem aumentar exponencialmente a quantidade de informação processada e disponibilizá-la em tempo real, elas, por si só, não são garante da qualidade com que ela é apresentada ao utilizador final. E é aqui que reside o grande desafio destas organizações, que procuram oferecer uma diversidade cada vez maior de produtos e serviços, obrigando-se, assim, a um esforço continuado de rigor, qualidade, inovação e aprendizagem.

A busca incessante da universalidade que nos é revelada pela história humana¹, a procura de uma ligação entre todos os homens do globo, pontuada por marcos simbólicos como foram o sistema métrico (instrumento fundamental de normalização a nível mundial) ou a criação de uma língua universal, aliados ao desenvolvimento e concretização de ideias como a de rede, nunca estiveram tão próximas de se realizar. Desde cedo metaforizado pelos naturalistas e anatomistas, a partir do séc. XIX o termo “rede” começa a aplicar-se a uma série de situações naturais e de infraestruturas criadas pelo Homem: rede hidrográfica, rede ferroviária (um dos mais eficazes meios de unificação e ligação territorial da era industrial), rede eléctrica e rede rodoviária, que se expandem em força no séc. XX, redes de telecomunicações que se iniciam com o telégrafo e o telefone e que são hoje a estrutura tecnológica essencial da comunicação a nível global. A ideia de rede aparece-nos, assim, como um poderoso meio estratégico² da globalização, através da qual se exploram energias, se amplia a circulação de pessoas e bens e se troca e difunde a informação.

¹ Um desenvolvimento desta ideia é feito por Armand Mattelart na sua obra *História da sociedade da informação* (Mattelart, 2002).

² A ideia de rede aliada ao conceito de estratégia remonta provavelmente, de acordo com Mattelart (2002), ao engenheiro militar Vauban (1633-1707) e ao seu pensamento sobre a organização do território, expresso no sistema de fortificações que concebeu e presente nas suas preocupações com a necessidade de facilitar a circulação de pessoas e mercadorias com vista à unificação das nações emergentes.

A exploração das redes permite hoje uma aproximação cada vez maior a um universo de saber alargado em que as bibliotecas e outros serviços de informação, actuando como mediadores entre o produtor/criador e o utilizador final, assumem um papel particularmente relevante. Nunca estiveram tão perto de concretização as ideias de Paul Otlet e de Henri Lafontaine sobre o “livro universal do saber” e o acesso universal ao conhecimento, ligadas à criação do Instituto Internacional de Bibliografia³ (que ambos fundaram em Bruxelas em 1895) e à tentativa de organização do conhecimento que empreenderam a partir da classificação decimal concebida por Melvil Dewey em 1876 e que deu origem à Classificação Decimal Universal. Paul Otlet (1868-1944), numa obra publicada em 1934⁴, antecipava mesmo a ideia de rede das redes que é hoje a Internet – uma rede universal de informação e documentação que ligaria “os centros produtores, distribuidores e utilizadores, de todas as especialidades e de toda a parte” (Mattelart, 2002, p. 51).

O caminho percorrido para este “acesso universal ao conhecimento” não é, contudo, isento de problemas e de contradições – se é verdade que hoje temos disponíveis meios tecnológicos e científicos que nos permitem pensar e agir “globalmente” (no sentido da “mundialização” de que falava Otlet) e que cada vez mais nos facilitam o acesso a todas as manifestações do conhecimento humano, sem limitações geográficas, é igualmente verdade que elas são inalcançáveis pela maioria da população mundial. As elevadas taxas de analfabetismo, o fosso entre países ricos e pobres e as diferenças de oportunidade inerentes a essa situação condicionam gravemente a concepção de “acesso livre e global”. A barreira linguística constitui outra limitação forte ao acesso, provocada por razões diversas: por um lado, a “imposição” da língua inglesa⁵ na maior parte dos recursos da Internet e em particular nas publicações científicas nas áreas das ciências ditas “naturais” e “exactas” (cuja difusão está muitas vezes condicionada a grupos de editores, sobretudo anglo-saxónicos, que se tornaram verdadeiros monopólios da edição científica), que

³ Organismo precursor da FID (*Fédération Internationale de Documentation*).

⁴ OTLET, Paul - *Traité de documentation : le livre sur le livre : théorie et pratique*. Bruxelles : Editions Mundaneum, 1934.

⁵ Como, no séc. XIX, a língua mais utilizada era a francesa, ou a alemã, no caso de domínios como a Química (UNESCO, 1971, p. 84).

acaba por favorecer a produção em língua inglesa em desfavor da produção científica expressa noutras línguas⁶; e, por outro lado, o facto de muitas das línguas em que estão expressos os conteúdos da Internet não serem compreendidas por todos os que podem aceder ao ciberespaço. A língua inglesa impõe-se, assim, como uma língua franca, um instrumento linguístico hoje necessário para aceder à informação; no entanto, existe uma ligação indissolúvel entre língua, pensamento e cultura e, de acordo com André Martinet (citado por Holzem, 1999, p. 39), a aprendizagem de uma língua estrangeira não significa atribuir novas palavras a objectos conhecidos, mas sim analisar de outra forma o que constitui o objecto da comunicação linguística; as línguas evoluem e aperfeiçoam-se com o uso, os vocabulários especializados contribuem para a sua vitalidade, cada língua contém em si uma série de escolhas sobre maneiras de representar o mundo, pelo que, ao marginalizar línguas, se está, na realidade, a marginalizar as culturas a que correspondem. Sem confundir “internacional” e “unilinguismo em inglês”, é necessário garantir um equilíbrio multilíngue na produção científica e, conseqüentemente, igualdade de oportunidades à sua divulgação⁷. Daí a importância de todas as iniciativas que visem permitir um acesso multilíngue à informação.

Antecedendo a extensão às bibliotecas do conceito de rede, a tentativa de possibilitar o acesso a grandes quantidades de informação com um esforço mínimo por parte do utilizador foi concretizada na criação de catálogos colectivos, iniciativas frequentemente desenvolvidas com forte participação das bibliotecas nacionais e que procuraram responder à necessidade de centralizar, num único local, a informação sobre os recursos (sintetizados em referências bibliográficas) existentes num maior ou menor número de bibliotecas, com o objectivo de responder às necessidades de uma pesquisa alargada de informação.

⁶ Veja-se, a este respeito, a tese defendida por Maryvonne Holzem em *Terminologie et documentation : pour une meilleure circulation des savoirs*. (Holzem, 1999).

⁷ Cf. Holzem, 1999, p.37-46.

Se os sistemas manuais tornavam difícil este tipo de iniciativa, dado o grande volume de referências e a necessidade, variável conforme os casos, de as multiplicar por vários catálogos (onomásticos, disdascálicos, sistemáticos...), a informatização das bibliotecas abriu caminho a que esta actividade se tornasse mais exequível, conduzindo à criação de bases de dados bibliográficos de tipo cooperativo, como foi o caso português da PORBASE – Base Nacional de Dados Bibliográficos. Os catálogos colectivos tornaram mais evidente a necessidade de normalização, já que os dados, embora com origens diferentes, tinham de apresentar um nível de identidade e consistência na descrição que possibilitasse a sua integração em bases comuns (suportadas por um determinado sistema informático) que respondessem de modo correcto, completo e preciso, às pesquisas de informação. O desenvolvimento das redes veio alargar ainda mais esta necessidade de normalização, estendendo-a da descrição de dados aos protocolos de comunicação e à capacidade de ligar e tornar compatíveis diferentes sistemas informáticos, desenvolvidos por diferentes organizações (empresariais ou não). E se as redes, que se expandem à escala planetária, são já resultado de um avançado nível de normalização, elas tornam igualmente indispensável um esforço continuado de cooperação que promova a elaboração de acordos e normas internacionais, a harmonização das normas nacionais desenvolvidas nos diversos países, assim como a sua revisão sistemática e constante à luz de novos desenvolvimentos e de novos condicionalismos.

Já no início da década de 80 do séc. XX a norma era apresentada como “um meio de comunicação posto à disposição de todos e que estreitam relações para além das fronteiras geográficas, entre os países.” (Moreira, 1980, p. 47-48). Ela é, de facto, um dos elementos-chave da interoperabilidade e está subjacente aos actuais sistemas de organização do conhecimento. Estes, construídos de acordo com estruturas racionais, lógicas, correspondem a uma das áreas que desde cedo se desenvolveu em bibliotecas e outros serviços de informação – a área da análise e da representação de assuntos. As possibilidades de acesso à informação não se restringem apenas aos dados descritivos e de localização dos recursos, elas estendem-se aos seus conteúdos, obrigando a uma abordagem que identifique as suas áreas

temáticas e os conceitos que veiculam. Este processo de análise sintetiza-se depois numa **linguagem documental** (linguagem convencional e controlada) através da qual se acede ao seu conteúdo intelectual. A linguagem documental funciona assim como mediadora entre o recurso e a possibilidade da sua descoberta e, para tal, deve obedecer a regras que evitem determinadas características da linguagem natural como, por exemplo, a sinonímia e a polissemia. As situações de ambiguidade, que o contexto da linguagem natural pode esclarecer, têm de ser controladas e limitadas na linguagem documental, sob pena de não permitirem uma pesquisa e recuperação correctas da informação pretendida. Além disso, têm de ser objecto de uma actualização permanente que tenha em consideração o desenvolvimento científico (e a evolução terminológica que o acompanha) e a própria dinâmica das línguas. Se, no caso das linguagens documentais de tipo notacional (como os sistemas de classificação) a questão do controlo terminológico não se tem colocado de modo premente e a sua estrutura é, frequentemente, de tipo hierárquico e categorial, já no caso das linguagens que utilizam termos retirados da linguagem natural para representar conceitos e assuntos se colocam problemas relacionados quer com a própria escolha dos termos, quer com as relações que entre eles se estabelecem. Daí que se tenham vindo a desenvolver instrumentos normativos que, ao criarem princípios e estruturas comuns de análise, representação e construção de linguagens documentais, procuram minorar os riscos da ambiguidade e da subjectividade, paralelamente à criação de normas com vocabulários especializados em várias áreas da indústria, da tecnologia, da documentação⁸ e da própria terminologia.

Contudo, apesar deste esforço normalizador e da elaboração de classificações e linguagens terminológicas largamente utilizadas a nível internacional, são

⁸ São exemplo de normas nesta área: NP 3715.1989, Documentação – *Método para a análise de documentos, determinação do seu conteúdo e selecção de termos de indexação*. Lisboa: IPQ; ISO 2788.1986 Documentation - *Guidelines for the establishment and development of monolingual thesauri*. [Genève]: ISO; ISO 999.1996 Information and documentation - *Guidelines for the content, organization and presentation of indexes*. [Genève]: ISO. ISO 5964.1985, Documentation – *Guidelines for the establishment and development of multilingual thesauri*. [Genève]: ISO; ISO 5127. 2001, Information and documentation – *Vocabulary*. [Genève]: ISO (em fase de adaptação a Norma Portuguesa).

ainda bastantes e variadas as linguagens documentais utilizados em sistemas de organização do conhecimento. É necessário que se conheçam as suas características, as semelhanças e as diferenças que apresentam entre si, para podermos compreender o tipo de desafio que se coloca ao acesso por assunto em ambientes de rede e a diversidade de problemas que têm de ser encarados e resolvidos.

As considerações que temos vindo a tecer levam a que, no presente trabalho, comecemos por abordar, no primeiro capítulo, a questão das **bases de dados bibliográficos** e dos **catálogos colectivos**, a **estrutura da informação** que contém, os **sistemas** que as suportam, os **aspectos normativos** que envolvem e os seus vários modelos, continuando o segundo capítulo com a análise de caso de uma base de dados bibliográfica de tipo cooperativo, a **PORBASE – Base Nacional de Dados Bibliográficos** e o terceiro com uma abordagem às questões relativas à **análise, representação e recuperação de assuntos**.

As duas últimas décadas do séc. XX e os primeiros anos do séc. XXI trouxeram a lume uma multiplicidade de ideias, iniciativas e projectos com o objectivo comum de contribuir para um mais fácil acesso à informação, utilizando os meios informáticos e as novas tecnologias disponíveis. Entre aqueles que têm vindo a desenvolver-se no âmbito do acesso por assunto seleccionámos, para objecto da análise desenvolvida no quarto capítulo, três projectos que, correspondendo a diferentes abordagens da questão, podem apontar caminhos que permitam desenvolver, numa nova perspectiva, os catálogos colectivos ou bases de dados cooperativas. São eles: o **projecto RENARDUS** – um exemplo de organização do conhecimento e de interoperabilidade entre diferentes sistemas de classificação e um modelo comum; o **projecto MACS** (*Multilingual Access to Subjects*) – um exemplo de interoperabilidade entre linguagens de indexação de diferentes bibliotecas nacionais, em diferentes línguas; e o **projecto HILT** (*High Level Thesaurus*) - uma proposta de interoperabilidade entre diferentes linguagens de indexação utilizadas por entidades tipologicamente distintas (bibliotecas, arquivos, museus). Procuraremos apresentar, no capítulo final, uma proposta de caminhos que podem ser

seguidos para melhorar o acesso à informação através da PORBASE – Base Nacional de Dados Bibliográficos.

2. Estrutura da informação em bases de dados cooperativas, ou catálogos colectivos

2.1 A norma ISO 2709 e os formatos MARC na génese das bases de dados bibliográficos

Como já referimos, o desenvolvimento da tecnologia informática, para além de otimizar as actividades e tarefas inerentes ao processamento bibliográfico e à recuperação da informação e de permitir a sua gestão integrada, conduziu à necessidade de aprofundamento do trabalho técnico a diversos níveis e, em particular, na área da normalização: com efeito, novos instrumentos normativos tiveram de ser criados para permitirem o tratamento informatizado da documentação e a troca de registos bibliográficos, ao mesmo tempo que se desenvolveram normas e formatos de descrição adequados à leitura por computador, como a norma ISO 2709⁹ e os formatos MARC, apenas para citar os exemplos mais óbvios no mundo biblioteconómico. A **norma ISO 2709** especifica os requisitos de um formato geral de troca de descrições bibliográficas e estabelece a estrutura dos registos informáticos para a comunicação entre sistemas de processamento de dados, isto é, para o intercâmbio de informação. O **formato MARC**, acrónimo de *MAchine Readable Cataloguing*, surgiu de uma iniciativa da Biblioteca do Congresso, em 1965-66, e permite aos computadores trocar, usar e interpretar a informação bibliográfica. A partir do início da década de 1970 começaram a surgir vários formatos da família MARC, relacionados com diferentes exigências e práticas catalográficas. O próprio formato MARC original evoluiu para o USMARC, na década de 80, e para o MARC 21, já na década de 90¹⁰. O formato UNIMARC foi criado pela IFLA (*International Federation of Library Associations and*

⁹ ISO 2709. 1996, Information et documentation – *Format pour l'échange d'information*. Genève : ISO, 6p.

¹⁰ Veja-se, sobre os formatos MARC, os endereços <URL: <http://www.ifla.org/VI/3/p1996-1/unimarc.htm>> [Consult. 27 Mar. 2003] e <URL: <http://www.loc.gov/marc/>>. Veja-se ainda Delsey (1996).

Institutions) para resolver a incompatibilidade existente entre os vários formatos MARC, permitindo a conversão de e para um formato universal, por forma a criar melhores condições para a troca de registos bibliográficos. Inicialmente previsto para a descrição bibliográfica de monografias e publicações em série, o formato foi-se alargando, ao longo dos anos, a todo o tipo de materiais, como o cartográfico, o musical ou os documentos electrónicos. A primeira edição de *UNIMARC – Universal MARC format*, da responsabilidade, como dissemos, da IFLA, data de 1977. Paralelamente ao UNIMARC / Bibliográfico têm vindo a desenvolver-se outros formatos UNIMARC, como o UNIMARC / Autoridades e, mais recentemente, um formato UNIMARC / Classificação e um formato UNIMARC / Existências. Do plano estratégico para a actividade nuclear do UNIMARC (*UNIMARC Core Activity*), desde o 2º trimestre de 2003 sediada em Portugal, na Biblioteca Nacional, consta a promoção de actividades de harmonização com o MARC 21 que garantam o objectivo primordial do UNIMARC – ser um formato universal de troca para o desenvolvimento do controlo bibliográfico universal¹¹.

Foi, pois, a necessidade de trocar informação bibliográfica e de desenvolver a cooperação tendente à constituição de catálogos colectivos, cujos dados deviam ser homogéneos, que obrigou à criação de um formato que pudesse ser interpretado por todos os sistemas das bibliotecas que o viessem a adoptar. (García Camarero e García Melero, 2001, p.43).

2.2 Os sistemas de suporte às bases de dados bibliográficos

O nascimento do formato MARC encontra-se ligado à fase inicial do processo de informatização das bibliotecas, a dos sistemas fechados, caracterizados pela existência de um computador central e de terminais que permitiam o processamento da informação à distância. Estas primeiras redes de computadores tinham uma estrutura hierárquica e um carácter homogéneo, quer a nível de equipamentos quer a nível de programas, aplicações e

¹¹ IFLA - *Strategic plan for the UNIMARC Core Activity*. IFLA, 2002. [Em linha]. [Consult. 9 Jun. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/VI/3/annual/spunimarc.htm>>

formatos. Seguiam um esquema de *anfitrião* ou *hóspede*, segundo o qual, para haver comunicação, o *hóspede* tinha de conhecer a linguagem do computador *anfitrião* e submeter-se às suas regras. O desenvolvimento da microinformática e a generalização do uso de microcomputadores permitiu uma alteração profunda nos sistemas e redes informáticas, com a criação dos chamados sistemas abertos e das redes baseadas no modelo *cliente – servidor*, no qual o *cliente* (computador usado pelo utilizador da rede para enviar ou receber mensagens) utiliza serviços oferecidos pelo *servidor* (computador que gere a transmissão de mensagens que circulam na rede e que foram pedidas ou enviadas por um computador cliente). Um sistema aberto, para além de comportar servidores e clientes, necessita ainda de canais de comunicação que pressupõem, para além da existência de uma “linha fixa”, o estabelecimento de acordos concretizáveis em protocolos, formatos e normas - uma vez que os serviços de informação cliente/servidor são independentes das respectivas plataformas informáticas, é o entendimento que cada computador possa ter de um protocolo que o pode tornar um cliente ou um servidor. É por isso que “os protocolos da Internet significaram uma transformação radical dos anteriores sistemas fechados, que dependiam da arquitectura dos computadores e da informação dos servidores” (García Camarero e García Melero, 2001, p.44).

Um dos primeiros protocolos de comunicação foi o **TCP** (*Transmission Control Program*), definido em 1974, tendo surgido mais tarde o protocolo **IP** (*Internet Protocol*), ambos protocolos gerais na arquitectura da Internet, versão pioneira (e simples) do modelo **OSI** (*Open Systems Interconnection*). Este modelo apresenta sete níveis: físico, conexão (correspondentes ao nível “sub-rede” da arquitectura da Internet), rede, transporte, sessão, apresentação e aplicação. O protocolo **IP** utiliza-se para especificar o sistema que identifica de forma única cada computador da rede¹², o que permite direccionar os dados ao mesmo tempo que possibilita a sua segmentação e “empacotamento” (como e quando devem ser divididas as mensagens para facilitar a sua transmissão). O protocolo **TCP** usa-se para transporte e controlo dos dados, que podem obedecer também a outros protocolos. O protocolo **IP** corresponde ao nível

¹² Cada um tem um número de identificação (o número IP) e uma designação (nome IP).

“rede” do modelo OSI e o protocolo TCP ao nível “transporte”. As aplicações correspondem aos níveis superiores do modelo (“sessão”, “apresentação” e “aplicação”). Nestas aplicações incluem-se os serviços de correio electrónico e os serviços de fornecimento de informação, de que se destacam os protocolos **FTP** (*File Transfer Protocol*) e **WAIS** (*Wide Area Information Servers*)¹³, este último uma ampliação do protocolo Z39.50 para pesquisa em texto integral, além da própria WWW (*World Wide Web*), que começou por usar a linguagem HTML (*Hypertext Markup Language*) e o protocolo **HTTP** (*Hypertext Transfer Protocol*), além de outras técnicas como o MIME (*Internet Media Type*), que especifica o tipo de dados de um ficheiro informático e o URL (*Uniform Resource Locator*), referência ou endereço de um recurso na Internet.

Entre os formatos de “escrita” e armazenamento dos documentos electrónicos de que hoje dispomos, encontra-se o mais antigo, geral e simples de todos, baseado na norma ASCII (*American Standard for Codification and Interchange of Information*), que define a codificação de um alfabeto, primeiro com 128 símbolos ou caracteres, mais tarde com 256, a que se seguiram ampliações sucessivas, de modo a cobrir o número crescente de aplicações alfabéticas, os sinais próprios de cada língua, os símbolos diacríticos e os símbolos científicos. Entre 1993 e 1997 a ISO (*International Standard Organisation*) publicou a norma 10646¹⁴, que corresponde ao *standard* UNICODE desenvolvido pelo UNICODE Consortium. A vantagem do formato ASCII residia principalmente no facto de poder ser lido e armazenado por qualquer computador, sem necessidade de um programa específico para aceder aos documentos que o utilizavam. Mas o desenvolvimento da edição electrónica, juntamente com o aparecimento da noção de “documento estruturado”, levou a que, logo a partir dos anos 60 do séc. XX, se comesçassem a desenvolver linguagens próprias para este novo tipo de documentos. É disso exemplo a GML (*Generalized*

¹³ Veja-se, sobre o WAIS e a sua utilização: NIEUWENHUYSEN, Paul (2004) - *Z39.50 and related protocols for search and retrieval*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.vub.ac.be/BIBLIO/nieuwenhuysen/courses/chapters/wais.pdf>>; SAINT PIERRE, Margaret [et al...] (1994) - *WAIS over Z39.50-1988*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.faqs.org/rfcs/rfc1625.html>>; e ARDÖ, Anders [et al.] ([1994]) - *Improving resource discovery and retrieval on the internet: the Nordic WAIS / World Wide Web Project: summary report*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.lub.lu.se/W4/summary.html>>.

¹⁴ ISO 10646: 1993-1997: *Information Technology. Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and basic multilingual plane*. Genève: ISO.

Markup Language), desenvolvida pela empresa IBM, concebido para que um mesmo ficheiro pudesse ser editado por diferentes aplicações com formatos de impressão distintos. À medida que os documentos se tornavam mais complexos e aumentava a sua difusão, foi-se sentindo a necessidade de um maior nível de normalização, surgindo assim um procedimento destinado a definir os tipos de documento e desse modo facilitar a edição e manipulação de ficheiros – o **DTD** (*Document Type Definition*). Já nos anos 80, a ANSI (*American National Standards Institute*) recorreu aos grupos que haviam desenvolvido a GML e, antes dela, o *GenCode*¹⁵, a fim de desenvolverem uma norma para a especificação, definição e uso de documentos com características de flexibilidade e portabilidade, trabalho de que resultou a **SGML** (*Standard Generalized Markup Language*), convertida em 1986 na norma ISO 8879. Embora facilitando a manipulação e a troca de documentos textuais, a SGML era uma linguagem complexa que não estava ainda suficientemente desenvolvida para documentos em rede nem se adequava aos primeiros passos da WWW, já na década de 90, na qual se viria a utilizar uma linguagem mais simples e acessível, a **HTML** (*Hypertext Markup Language*), com elementos da SGML mas com outras possibilidades que permitiam operações de hipertexto¹⁶. A HTML, de que saíram até hoje várias versões¹⁷, corresponde a “um conjunto de códigos ASCII usados em documentos na Web para lhes dar a forma de hipertexto/hipermédia” (García Camarero e García Melero, 2001, p. 69). Mas a HTML não possuía algumas das características importantes da SGML: a extensibilidade¹⁸, a estrutura e a validação. Tal facto conduziu a uma nova linguagem de marcação para definir documentos estruturados, a **XML** (*eXtensible Markup Language*), subconjunto da SGML, que não só mantém as suas propriedades de extensibilidade, estrutura e validação como também responde às necessidades da WWW, já que se trata de um formato normalizado destinado especificamente à transmissão de dados

¹⁵ Um dos primeiros códigos para a representação de textos estruturados (García Camarero e García Melero, 2001, p. 68).

¹⁶ A concepção de hipertexto começou a surgir no pós-guerra e foi desenvolvido nos anos 60. O documento hipertextual permite-nos circular de múltiplas formas entre os elementos que o constituem (Guédon, 2001, p. 52).

¹⁷ A versão 4.0 (1997) adoptou a Norma ISO/IEC 10646, correspondente ao UNICODE, para os conjuntos de caracteres HTML.

¹⁸ Extensibilidade: “característica de uma linguagem que consiste em mecanismos de extensão que permitem a definição de novos elementos em termos do conjunto de regras que constituem a axiomática da linguagem” (Sousa, 2001, p. 104).

estruturados às aplicações *Web*. Supera assim as limitações da HTML e permite a interoperabilidade tanto com esta como com a SGML.

Considerados no âmbito da informação existente na *Web*, os metadados descobrem e descrevem os recursos nela disponíveis, utilizando linguagens de marcação como a HTML ou a XML, atrás referidas. Para o fazerem de um modo completo e fiável, estão organizados de acordo com tipologias – metadados descritivos, administrativos e estruturais, respectivamente com a finalidade de descrever o recurso (à semelhança do que faz a descrição bibliográfica), de informar sobre aspectos legais, formatos, resolução de imagem, preservação ou persistência do recurso, entre outros, e sobre a própria estrutura interna do recurso, como paginação, capítulos, ilustrações, gráficos, etc. (Campos, 2002, p.4).

A necessidade de normalização em relação aos metadados levou a que a ISO e o CEN (*Comité Européen de Normalisation*) reconhecessem a importância de normalizar a estrutura criada no âmbito do DCMI (*Dublin Core Metadata Initiative*) e genericamente conhecida por *Dublin Core*, ou *Dublin Core Metadata Element Set* (Campos, 2002, p.5). Este núcleo básico de elementos de metadados foi elaborado ao longo de várias reuniões de trabalho em que participaram especialistas de bibliotecas, de redes e investigadores sobre bibliotecas digitais¹⁹, a primeira das quais se realizou em Dublin (Ohio, E.U.A.) em 1995, resultando assim de um consenso internacional e interdisciplinar. Para García Camarero e García Melero (2001, p.152-153), entre as características que tornam o *Dublin Core* adequado à catalogação de documentos electrónicos destacam-se, além do já referido consenso internacional, a simplicidade, a interoperabilidade semântica, a flexibilidade e extensibilidade, a modularidade e a independência sintáctica, que o torna passível de ser incorporado em várias sintaxes já existentes, como os formatos bibliográficos normalizados – daí que tenham vindo a desenvolver-se, nos últimos anos, iniciativas de mapeamento entre o *Dublin Core* e os formatos MARC.

¹⁹ Participaram mais de 20 países, de quatro continentes (García Camarero e García Melero, 2001, p. 152).

Tendo os metadados, entre as suas várias funções, a possibilidade de informar sobre o valor relativo dos conteúdos recuperados, o *WWW Consortium* (W3C) começou a desenvolver, em 1995, um mecanismo de qualificação dos resultados de uma pesquisa, com o qual seria possível indicar ao utilizador o valor atribuído a cada conteúdo seleccionado em relação à pesquisa por ele efectuada. Nasceu assim a PICS (*Platform for Internet Content Selection*)²⁰, cujos perfis permitiam igualmente aos utilizadores seleccionarem a informação que desejavam receber através da filtragem de conteúdos, baseada nas regras básicas de selecção “rejeitar” e “aceitar”. As condições de aceitação ou rejeição podiam ser um conjunto de URLs, ou determinados valores de categorias indicadas.

O desenvolvimento da PICS, por um lado, e a heterogeneidade dos objectos digitais e dos metadados que os descrevem, por outro, revelou a necessidade de se desenvolver uma infraestrutura que permitisse “a codificação, o intercâmbio e a reestruturação de metadados” e tornasse possível “a interoperabilidade dos metadados mediante mecanismos apoiados em convenções semânticas, sintácticas e estruturais genericamente aceites” (García Camarero e García Melero, 2001, p. 172). Nascia assim, também no âmbito do W3C, a *RDF (Resource Description Framework)*²¹. Começando por ser uma extensão da tecnologia de descrição de conteúdo da PICS e utilizando a sintaxe da linguagem XML como sintaxe de intercâmbio e processamento de metadados, este “formato universal de dados existentes na Web” permite, usando um simples modelo relacional, “combinar, exportar e partilhar dados estruturados e semi-estruturados através de diferentes aplicações” (Berners-Lee et al., 2004). Qualquer recurso que possua um identificador normalizado (como o URI) é susceptível de ser descrito segundo o modelo RDF. Este baseia-se no facto de os recursos possuírem propriedades, atributos ou características e de aos tipos de propriedades corresponderem valores, sendo a descrição um conjunto de propriedades que se referem ao mesmo recurso. Trata-se de um modelo que representa os recursos e as suas descrições de

²⁰ <URL: <http://www.w3.org/PICS/>>

²¹ <URL: <http://www.w3.org/RDF/>>

forma independente da sintaxe e que, através dos três tipos de elementos (recursos, tipos de propriedades e valores correspondentes), pode expressar a semântica numa codificação legível por máquina, sem ambiguidades. Além disso, possui um modelo de dados para descrição de outras descrições, que pode usar-se, por exemplo, para certificar a credibilidade de uma determinada descrição, ou na descrição de colecções de recursos (García Camarero e García Melero, 2001, p. 172-174). As características deste formato de dados e a sua aplicabilidade aos objectos digitais tornam-no um dos pilares da *Web* semântica – termo que designa “uma extensão da *Web*, na qual a informação adquire um significado claramente definido, melhorando o trabalho cooperativo de computadores e pessoas” (Berners-Lee, Hendler e Lassila, 2001) e “uma visão para o futuro da *Web* segundo a qual a informação adquire um significado explícito, tornando mais fácil o processamento e integração automáticos da informação disponível na *Web*” (McGuinness e Harmelen, 2004, [p.3]).

Se nos debruçámos sobre as características que têm vindo a desenhar a evolução dos sistemas abertos e da própria *WWW*, é por considerarmos que elas nos ajudam a compreender a evolução dos sistemas utilizados pelas bibliotecas nas últimas décadas, a sua adequação às novas realidades de produção, circulação, armazenamento e difusão da informação, assim como os esforços desenvolvidos para adaptação das estruturas institucionais a essas realidades.

2.3 Cooperação e normalização em bases de dados: o registo bibliográfico e o registo de autoridade

Paralelamente aos desenvolvimentos que acabámos de referir, abria-se caminho a uma série de novas iniciativas no âmbito da cooperação. Como já referimos, o esforço de normalização acompanhou desde sempre o desenvolvimento das bases de dados bibliográficos de tipo cooperativo (os chamados catálogos colectivos informatizados, quase sempre em linha), já que, para uma eficaz recuperação da informação, se impunham critérios comuns, quer a nível da descrição bibliográfica, quer a nível das formas de

entrada dos nomes de autores, fossem eles pessoas individuais ou colectivas, quer ainda a nível da representação dos assuntos. Assim, com o objectivo de possibilitar uma efectiva recuperação da informação desenvolveram-se, paralelamente às bases bibliográficas e interagindo com elas, as bases de autoridades, obedecendo também estas a regras e a formatos uniformizados. Estas bases de autoridades permitem a pesquisa e recuperação da informação referente a um determinado autor (pessoa física ou colectividade) ou a um determinado assunto, independentemente da forma como o pedido é formulado: isto é, quer este tenha sido expressado através de um pseudónimo, de um nome real ou de outra variante do nome, no caso dos autores pessoa física, ou de um sinónimo ou quase sinónimo (termo “rejeitado”), no caso dos assuntos. A estrutura de um registo de autoridade possibilita o estabelecimento de uma série de relações e de referências remissivas que, correctamente preenchidas, facilitam a gestão integrada dos ficheiros de autoridade e bibliográfico de modo a satisfazer da melhor forma os pedidos de informação (ver anexo 1 - registos de autoridade (UNIMARC / Autoridades)).

A importância das bases de dados de autoridade tem sido amplamente reconhecida a nível internacional, dando origem a vários debates, projectos e iniciativas no âmbito da normalização, da melhoria do acesso e da partilha dos registos criados nos diferentes países. Os esforços para a criação de instrumentos normativos que permitam a consistência das formas dos cabeçalhos, essenciais ao controlo de autoridade, radicam nos Princípios de Paris (1961), elaborados num contexto ainda longínquo dos desafios e oportunidades que a Internet viria a potenciar. Um projecto apoiado pela Comunidade Europeia e que se desenvolveu entre 1995 e 1997 – o projecto AUTHOR – veio demonstrar que as práticas e as regras referentes à criação de registos de autoridade eram muito diferentes nos vários países envolvidos (Reino Unido, França, Bélgica, Espanha e Portugal), embora todos seguissem os Princípios de Paris. Contudo, as diferenças manifestavam-se sobretudo a nível de indicativos geográficos e de elementos de identificação ou distinção, dos esquemas de transliteração e romanização, dos formatos MARC utilizados, na grafia (numa mesma língua mas de países diferentes), dos usos linguísticos relativos aos nomes de autores clássicos ou personalidades com

denominações consagradas pelo uso (Campos, 2003, p. 9-11), daí que instituições como a IFLA tenham encetado a revisão dos princípios estabelecidos e desenvolvido novas directrizes numa tentativa de ultrapassar as dificuldades detectadas. Produziram-se assim os *Guidelines for Authority Records and References* (GARR)²², publicados em 2001, e encetou-se a revisão dos Princípios de Paris à luz da nova realidade tecnológica e dos novos conceitos veiculados pelos *Functional Requirements for Bibliographic Records* (FRBR)²³ (Campos, 2003, p.14).

Paralelamente têm vindo a desenvolver-se vários projectos nesta área, como o projecto LEAF (*Linking and Exploring Authority Files*), financiado pelo programa “Tecnologias da Sociedade de Informação” do 5º Programa Quadro da Comissão Europeia, que decorreu entre 2001 e 2004 e tinha como objectivo conceber um modelo de arquitectura para o estabelecimento de ligações entre registos de autoridade distribuídos e permitir o seu acesso. As informações recuperadas como resultados de pesquisas seriam armazenadas num ficheiro central de autoridades que cresceria assim à medida que aquelas fossem sendo realizadas, reflectindo desse modo os dados mais relevantes para os utilizadores. Os resultados do projecto seriam implementados na expansão da rede internacional de serviços de pesquisa e recuperação em linha resultante do projecto MALVINE (*Manuscript and Letters via Integrated Networks in Europe*). A Biblioteca Nacional de Portugal participou nos dois projectos²⁴.

No sentido de tornar efectiva a criação de um verdadeiro sistema internacional de autoridades, cuja necessidade foi formulada em 1978, por ocasião da Conferência da IFLA realizada na então Checoslováquia, paralelamente à decisão de desenvolver o formato UNIMARC para as formas de autoridade e de conceber métodos para uma troca eficiente e eficaz dos dados de autoridade (Plassard, 2003) e na sequência dos trabalhos desde então

²² Revisão dos GARE – *Guidelines for Authority and Reference Entries*, de 1984.

²³ IFLA. IFLA Study Group on the Functional Requirements for Bibliographic Records - *Functional Requirements for Bibliographic Records : final report*. München : K. G. Saur, 1998.

²⁴ Veja-se, sobre o projecto MALVINE e a participação da Biblioteca Nacional, OLIVEIRA, António Braz de, VASCONCELOS, Manuela; FREIRE, Nuno (2000) - O Projecto MALVINE: manuscritos literários em rede europeia, através do protocolo Z39.50. In JORNADAS PORBASE, 8, Lisboa, 2000 – *Sistemas de informação em bibliotecas do novo milénio*. Lisboa: Biblioteca Nacional, 2000.

desenvolvidos surgiu, em 1996, no seio do UBCIM (*Universal Bibliographic Control and International MARC Core Activity*)²⁵, um grupo de trabalho cujo objectivo era analisar o sistema internacional de autoridades e criar um ISADN – *International Standard Authority Data Number*. Em 1998, a *Deutsche Bibliothek* (Biblioteca Alemã) e a Biblioteca do Congresso conceberam um novo projecto, o *Virtual International Authority File* (VIAF), cujo objectivo era uma redução efectiva dos custos de catalogação tornando acessíveis, a nível mundial, os registos de autoridade. A aplicação do controlo de autoridade ao ambiente *web* estenderia assim a este domínio as vantagens da precisão nas pesquisas, para além de permitir a navegação através da estrutura de referências e uma capacidade alargada para estabelecer ligações (Tillet, 2002, p. 118). Em 2003, a OCLC (*Online Computer Library Center, Inc.*) juntou-se aos parceiros iniciais e foi assinado um acordo²⁶ com o objectivo de desenvolver o VIAF combinando os ficheiros de autoridade de nome de pessoa física da Biblioteca do Congresso e da Biblioteca Alemã num único serviço de autoridade (nome de pessoa) e tornando-o acessível através de um servidor OAI (*Open Archive Initiative*)²⁷. A longo prazo, o projecto VIAF incluirá as autoridades de várias bibliotecas nacionais num serviço global que deverá ser acessível aos utilizadores a nível mundial através da *Web*. Segundo Barbara Tillet, responsável pela política de catalogação da Biblioteca do Congresso, “o projecto VIAF pode ser encarado como um dos blocos básicos da construção da ‘*Web Semântica*’. O utilizador final poderá pesquisar registos existentes em todo o mundo, com o seu sistema de escrita e na sua língua, usando as formas com que estão familiarizados.”²⁸

²⁵ O UBCIM encerrou a sua actividade no seio da IFLA em Março de 2003, porém as actividades de controlo bibliográfico continuam, através dos grupos de trabalho constituídos para esse efeito e sob responsabilidade do ICABS (*IFLA / CDNL Alliance for Bibliographic Standards*). Ver <URL: <http://www.ifla.org/VI/7/icabs.htm>>.

²⁶ Este acordo foi assinado durante o Congresso Internacional da IFLA realizado em Berlim.

²⁷ A *Open Archives Initiative* – OAI tem como objectivo permitir o acesso a material existente na *web* através de repositórios interoperativos para partilha de metadados, edição e arquivo. O OAI – *Protocol for Metadata Harvesting* (OAI-PMH) define um mecanismo de recolha de registos a partir de repositórios de informação. Não se destina a efectuar pesquisas cruzadas, mas sim a recolher e colocar os dados num único local (“*harvesting*”). Veja-se, para detalhes sobre a OAI, <URL: <http://www.oaforum.org/>>.

²⁸ Citada em: OCLC, Die Deutsche Bibliothek and the Library of Congress to develop Virtual International Authority File. *OCLC Abstracts*. Vol. 6, n.º 35 (Set. 2003). [Em linha]. [Consult. 16 Set. 2003]. Disponível na WWW:

<URL: <http://www5.oclc.org/downloads/design/abstracts/09022003/viaf.htm>>.

Todos estes trabalhos reconhecem e confirmam, na realidade, as afirmações de Michael Gorman de que “o controlo bibliográfico é literalmente impossível sem o *controlo de autoridade*. A catalogação não pode existir sem pontos de acesso normalizados e o *controlo de autoridade* é o mecanismo mediante o qual obtemos o nível necessário de normalização” (Gorman, 2003, p.1).

2.4 Características gerais das bases de dados cooperativas ou catálogos colectivos

Quando nos referimos a catálogos colectivos no âmbito das bibliotecas podemos estar a aludir a realidades que, tendo em comum o facto de reunirem informação bibliográfica existente em várias instituições, podem contudo apresentar diferenças. Com efeito, os catálogos colectivos podem ser **especializados**, podem organizar-se por **tipologia documental** (de publicações periódicas, por exemplo)²⁹ e podem ser **nacionais**, **internacionais**, de **âmbito regional** ou concebidos de acordo com outros critérios de ordem geográfica³⁰. Podem, em suma, constituir-se segundo a vontade e os objectivos definidos por aqueles que os criam e que neles participam. Se exceptuarmos o caso específico daqueles que se especializam numa única área temática, a multidisciplinaridade é, em geral, uma das características dos catálogos colectivos. Reunindo informação oriunda de bibliotecas de diferentes tipos e âmbitos temáticos, espelham igualmente, além de diferentes critérios de análise, diferentes modelos de organização e representação dos assuntos. Estes vão dos sistemas de classificação de âmbito universal às linguagens terminológicas especializadas, passando por esquemas de classificação especializados e por linguagens terminológicas de carácter mais geral. É, de facto, grande a diversidade de sistemas de

²⁹ Como o Catálogo Colectivo Alemão de Publicações em Série (ZDB), gerido pela *Staatbibliothek zu Berlin* e disponibilizado num servidor da *Deutsche Bibliothek*, ou o Catálogo Colectivo Suíço das Publicações em Série que, a partir de meados de 2002, deu origem ao Portal Suíço dos Periódicos (PSP), acessível através do endereço <URL: http://www.swiss-serials.ch/index_fr.html.>

³⁰ Estão neste último caso, também a título de exemplo, o Catálogo Colectivo do Consórcio de Bibliotecas do Sudoeste da Alemanha (SWB - *Südwestdeutscher Bibliothekverbund*), desenvolvido e mantido pelo *Bibliotheksservice-Zentrum* (BSZ) e as redes de bibliotecas suíças RERO (ReseauROmand), da Suíça Ocidental e Sbt (Sistema Bibliotecario Ticinese), que coordena o catálogo cantonal e o catálogo escolástico.

tratamento de assuntos e de linguagens documentais que se propõem representá-los, tema que mais adiante retomaremos.

Outra das características que, hoje em dia, têm os catálogos colectivos, é o facto de poderem reunir descrições de informação disponível numa grande diversidade de suportes. Se é verdade que, para além de documentação predominantemente escrita, as bibliotecas desde cedo reuniram outro tipo de documentos, como os cartográficos, os iconográficos ou os musicais e, mais recentemente, os documentos audiovisuais e os documentos multimédia, são os documentos electrónicos que estão hoje no cerne do interesse (e também das preocupações) das bibliotecas, não tanto pelas questões inerentes à sua descrição, pois já existem instrumentos normativos e formatos de catalogação para esse tipo de documento, mas pelo que representam em si de novidade estrutural em relação ao tipo de informação anteriormente veiculada por um suporte físico, editada nos circuitos comerciais ou produzida e distribuída no âmbito de instituições, colectividades eventuais ou como resultado de trabalhos de investigação individual ou colectiva (geralmente sob a forma de *literatura cinzenta*³¹), levantando novos problemas a nível da recuperação.

Foi a partir do advento da Internet e da sua rápida expansão com a *World Wide Web*³² que as instituições tradicionais de selecção, tratamento e difusão da informação começaram a colocar a si próprias novas questões inerentes, precisamente, à selecção, tratamento e difusão da informação veiculada pela rede das redes. Para além da oportunidade de alargarem ao nível do globo a divulgação dos seus catálogos, as bibliotecas viram desde logo que um novo caminho se lhes abria enquanto mediadoras entre os recursos informativos disponíveis na Internet e os seus utilizadores. Mas como seleccionar e organizar a informação contida nesse gigantesco acervo? E como torná-la

³¹ Considera-se *literatura cinzenta* "o material produzido à margem dos circuitos comerciais de publicação, de distribuição, de controlo bibliográfico ou de aquisição por livreiros ou agentes de assinaturas." (Cf. IGLWG 1995, cit. por CORREIA, Ana Maria Ramalho e NETO, Miguel de Castro – *LIZA – Repositório de literatura científica cinzenta, produzida em Portugal : relatório final*, p. 1).

³² "Sistema hipermédia distribuído desenvolvido no CERN (*European Organisation for Nuclear Research*) a partir de 1989, baseado numa arquitectura cliente/servidor e que permite disponibilizar documentos hipertexto. A convivialidade dos programas cliente WWW, como o Mosaic, transformou a Web num utensílio universal de acesso aos recursos Internet" (Guédon, 2001, p. 121).

acessível a um número sempre crescente, cada vez mais heterogéneo (e também mais desconhecido) de utilizadores?

A heterogeneidade é uma característica fundamental da informação existente na rede, assim como a sua volatilidade. Heterogeneidade a nível de conteúdos, a nível dos seus produtores, a nível linguístico. E também a nível da própria organização da informação: podemos encontrar na rede informação secundária fortemente estruturada, como os casos dos catálogos individuais de bibliotecas ou dos catálogos colectivos, informação secundária pouco estruturada como, por exemplo, a que consta de alguns directórios³³ e, ainda, todo um manancial de informação primária não tratada. Esta heterogeneidade levanta, obviamente, problemas a nível da recuperação: se, por um lado, a informação tratada e estruturada em catálogos bibliográficos já é reflexo de diferentes opções de descrição e de diferentes sistemas e linguagens documentais de representação de assuntos, por outro lado a informação não incluída em catálogos bibliográficos, seja ela secundária e pouco estruturada ou primária, só pode ser recuperada, em parte, através dos motores de pesquisa e da própria navegação na rede. Coloca-se, assim, um duplo desafio: o de pôr todos os diferentes sistemas a dialogar, estabelecer parâmetros comuns e esquemas de equivalência (mapeamentos) para criar formas de acesso à informação sem necessidade de pesquisar individualmente cada catálogo, por um lado; e providenciar a selecção e organização da informação, por outro.

Ao mesmo tempo que transformaram profundamente as formas de produção e distribuição da informação, dando origem a um mundo de recursos que todos os dias se modifica na ordem dos milhões de referências (entre as que são publicadas pela primeira vez, as já existentes que se actualizam e as que são apagadas), a Internet e a WWW (*World Wide Web*) vieram simultaneamente possibilitar e criar a necessidade de novos instrumentos tecnológicos e normativos, novas metodologias e novos procedimentos que permitissem o acesso a toda essa "informação nova" e dinâmica, o seu armazenamento e a sua preservação em termos de futuro. Como atrás referimos, as bases de dados bibliográficas, com tradições de normalização e de resolução de

³³ Ainda que alguns deles sejam também bastante estruturados, à semelhança dos catálogos.

problemas de incompatibilidade, têm, desde então, entre outros desafios, o de lidar com uma nova realidade, a do documento electrónico, que coloca novos problemas a nível da selecção, da descrição, do armazenamento e da recuperação. Esta já não é feita apenas em termos de obtenção de referências bibliográficas, mas igualmente em termos de acesso ao próprio documento, que, em linhas gerais, é facilmente manipulável, susceptível de possuir várias ligações, transformável em diversos formatos, preparado para a pesquisa de conteúdos, imediatamente transportável por meio de redes de trabalho e passível de distribuição ilimitada (Garcia Jimenez, 1999, p.137). Na realidade, para além do seu conteúdo, os documentos electrónicos (sejam eles resultado da digitalização de livros, manuscritos, imagens ou vídeos, por exemplo, ou “geneticamente digitais”) possuem outra informação estrutural, técnica, necessária ao seu tratamento e gestão; por essa razão muitos os designam por *objectos digitais*, por considerarem que esta denominação se adequa melhor do que o termo *documento* ao conjunto de todos os aspectos de que se revestem (García Camarero e García Melero, 2001, p.144).

Torna-se assim necessário definir novos modelos de recuperação, adequados à navegação que a rede permite e dela característicos, atendendo sempre ao facto de a *Web* ser “um espaço onde coexistem dois mundos diferentes, com duas culturas distintas: a cultura das bases de dados e a cultura da tecnologia *Web*” (Cordeiro, 2001, p.6); dois mundos diferentes a nível técnico e a nível conceptual – as bases de dados são mais consistentes, mantêm a sua integridade e permanência, possuem serviços de pesquisa eficientes e os seus recursos são mais estruturados, ao passo que na *Web* não existe a garantia de integridade e permanência dos recursos, o acesso a estes depende de serviços externos e, embora os conjuntos de recursos sejam ilimitados, são igualmente pouco estruturados (Cordeiro, 2001, p.6-7). Há ainda a considerar o manancial de informação existente na *Web* mas que os motores de pesquisa gerais não incluem na sua colecção de páginas, como é o caso da chamada “*Web invisível*”.

Rapidamente se percebeu que novas oportunidades, mas também novas exigências, surgiam a nível da cooperação. O desenvolvimento dos sistemas

abertos e da tecnologia do âmbito da Internet oferece aos catálogos colectivos a possibilidade de se estruturarem de diferentes formas, podendo assumir um carácter menos centralizador, mais aberto e mais partilhado, com o objectivo de melhorar o acesso à informação; mas uma tal reestruturação implica transformações por vezes profundas na forma de trabalhar das organizações.

2.5 Os catálogos colectivos e os seus modelos

Os catálogos colectivos, seja qual for o modelo que seguem, têm como funções principais organizar e difundir o conhecimento, preservando e armazenando a informação de modo a que esta possa ser eficazmente recuperada com os melhores níveis de revocação e precisão. São precisamente estes os dois conceitos essenciais que, segundo Matthew J. Dovey (2000), ilustram a confiança que um utilizador pode ter na pesquisa que efectua: a obtenção de toda a informação pretendida, independentemente da forma do(s) termo(s) usado(s) na formulação da pesquisa (o que pressupõe uma estrutura subjacente de referências cruzadas), no primeiro caso, e a garantia de que os resultados da pesquisa são de facto relevantes, no segundo.

Existem basicamente, segundo Matthew J. Dovey (2000), dois grandes modelos de catálogos colectivos³⁴: os modelos físicos, que podemos também designar por catálogos colectivos centralizados / concretos (Campos, 1995); e os modelos virtuais, ou catálogos colectivos virtuais (Campos, 1995). Segundo Dovey (2000), no grupo dos modelos físicos podemos identificar quatro tipos diferentes, de acordo com a forma como se encontram organizados:

- o catálogo colectivo é o catálogo principal para as bibliotecas que nele participam;
- o catálogo colectivo recebe registos exportados pelos catálogos locais;
- os registos são catalogados no catálogo central e importados pelos catálogos locais;
- os catálogos locais e o catálogo central são actualizados em simultâneo, o que pressupõe uma catalogação distribuída.

³⁴ Dovey identifica ainda um possível terceiro modelo, o dos catálogos múltiplos, indexados localmente mas também a nível central, sendo nestes índices que a pesquisa é realizada, embora os registos que aparecem nos resultados venham directamente dos catálogos locais.

Quanto aos catálogos virtuais, em que se incluem os denominados "clumps"³⁵, visam permitir ao utilizador, através de um único interface, efectuar a pesquisa a uma série de catálogos distribuídos como se de um catálogo único se tratasse. Um exemplo de catálogos que optaram por esta via é o CAIRNS (*Co-operative Information Retrieval Network for Scotland*)³⁶, um catálogo colectivo distribuído, baseado em Z39.50, para as bibliotecas escocesas. Neste, os "miniclumps" são conjuntos de colecções pré-definidos para grupos especiais de utilizadores (nos quais se incluem projectos e parcerias de bibliotecas).

Tem sido reconhecido que os catálogos físicos permitem uma maior consistência, já que um controlo centralizado zela pelo cumprimento das regras e normas estabelecidas e procura evitar situações como, por exemplo, a duplicação de registos ou a incoerência em relação às entradas de autoridade e à representação dos assuntos, permitindo assim a obtenção de melhores resultados a nível da recuperação. No entanto, é igualmente aceite que a manutenção deste tipo de catálogos é pesada, pois requer a afectação de recursos bastante significativos por parte, sobretudo, da entidade responsável pela sua gestão. Por outro lado, o crescimento destes catálogos depende do ritmo de envio dos registos, por parte das bibliotecas cooperantes, para a base central e do trabalho executado sobre esses registos, por parte da entidade gestora, o que provoca, frequentemente, dificuldades em manter um nível aceitável de actualidade do conteúdo da base. Este problema não se coloca no caso dos catálogos virtuais pois, uma vez que não existe um catálogo único, central, que recebe os registos mas sim uma rede de catálogos locais que podem comunicar entre si, a informação obtida através das pesquisas será sempre aquela que nesse momento existe nas respectivas bases de dados. A

³⁵ Um "clump" pode definir-se "um grupo de catálogos distribuídos, através dos quais se pode pesquisar via Z39.50 e de outros meios que assegurem a interoperabilidade entre os catálogos" (Cf. *CAIRNS final report : an embryonic cross-sectorial, cross-domain national networked information service for Scotland?*. Maio 2001. [Em linha]. [Consult. 6 Dez. 2004]. Disponível na WWW: <URL: <http://cairns.lib.gla.ac.uk/cairnsfinal.pdf>>).

³⁶ CAIRNS – Co-operAtive Information Retrieval Network for Scotland. Trata-se de um serviço de informação em rede escocês, ou "Portal escocês", desenvolvido a partir de um projecto de investigação com o mesmo nome, financiado pela e-Lib e apoiado pela Scottish Confederation of University and Research Libraries (SCURL), que terminou no final de Dezembro de 2000 (Cf. *CAIRNS final report*) Para mais informações sobre o CAIRNS, veja-se <URL: <http://cairns.lib.strath.ac.uk/>>.

responsabilidade de manutenção de um catálogo virtual está, deste modo, muito mais distribuída por todos aqueles que nele participam. Contudo, para garantirem bons níveis de revocação e precisão, os catálogos virtuais colocam outro tipo de questões, que são aquelas que se prendem com a interoperabilidade. A **interoperabilidade** é, hoje em dia, um conceito essencial, uma das “palavras-chave” da gestão da informação. William E. Moen (2003) define-a a dois níveis: a nível do sistema e a nível do utilizador. No primeiro caso, considera-a “a capacidade de dois ou mais sistemas ou componentes trocarem informação e utilizarem-na sem esforço especial por parte de qualquer dos sistemas”. No segundo caso, a interoperabilidade é definida como a “capacidade que o utilizador tem de pesquisar e recuperar com sucesso ou de interagir com a informação de uma forma significativa e confiando nos resultados”, ou ainda “a condição que se atinge quando dois ou mais sistemas técnicos podem trocar informação directamente, de tal modo que satisfaça os utilizadores do sistema”. Moen identifica ainda os principais factores que afectam a interoperabilidade: a existência de múltiplos protocolos e de múltiplos formatos de dados, sintaxes e esquemas de metadados, de múltiplos vocabulários, ontologias e disciplinas, de múltiplas linguagens e conjuntos de caracteres, de múltiplos sistemas e de diversas políticas de indexação, de normalização e extracção de palavras. Como facilmente se compreende, a interoperabilidade não pode ser encarada apenas numa perspectiva meramente tecnológica - é mais do que compatibilizar *hardware* e *software* (Miller, 2000). Com efeito, a interoperabilidade existe quando sistemas, procedimentos e cultura de uma organização são geridos de forma a maximizar oportunidades para troca e reutilização de informação (Miller, 2000). Se a uma vertente tecnológica que possibilite uma comunicação fácil e normalizada entre sistemas informáticos se juntarem “factores de gestão de conteúdos”, como a utilização de um ficheiro de autoridade de nome de pessoa e de colectividade ou um mesmo “instrumento de controlo das formas de acesso por assunto”, a interoperabilidade será mais conseguida (Cordeiro, 2001, p.5). Além disso, ela assenta necessariamente em acordos, quer de âmbito normativo, quer a nível de procedimentos, o que pressupõe uma avaliação das necessidades globais e locais no ambiente de rede e a vontade, por parte dos que formam a comunidade dos serviços de informação, de uma real cooperação relativa aos

objectivos que se pretendem atingir (Moen, 2003), o que pode implicar mudanças radicais nas formas como as organizações trabalham e, em especial, nas suas atitudes em relação à informação (Miller, 2000).

Por todas estas razões, o desafio dos catálogos virtuais não é fácil e a atestar a sua complexidade estão os projectos e trabalhos que se têm vindo a desenvolver, com mais acuidade desde a última década, sobre diversos aspectos que lhes estão subjacentes e que passam pela criação de modelos de metadados e mapeamentos entre formatos, sempre com o objectivo final de permitir aos utilizadores o acesso mais fácil e imediato a toda a informação existente nesses múltiplos catálogos que constituem o "catálogo virtual". O desenvolvimento de protocolos de rede como o Z39.50³⁷, que permite o diálogo entre bases de dados bibliográficos usando um conjunto de regras definido *a priori*, só consegue atingir os seus objectivos de interoperabilidade se os sistemas seguirem, na sua arquitectura e estrutura de dados, algumas normas e regras comuns que permitam gerar concordâncias e equivalências. É essa a razão pela qual, no caso do protocolo Z39.50, a sua aplicação se verifica sobretudo em sistemas de informação bibliográfica, cujas arquitectura e estrutura de dados assentam em bases normativas bem definidas, quer a nível dos "tipos e níveis das entidades lógicas de informação", quer a nível dos "modelos de organização dos elementos de informação, por exemplo ISBD e códigos de catalogação", quer ainda a nível "sintáctico/semântico dos formatos de representação dos elementos de dados (os formatos legíveis por computador, nomeadamente os formatos MARC e o seu modelo comum, o ISO 2709)" (Cordeiro, 2001, p.5). As especificações do Z39.50, como o *Bath Profile*³⁸, vieram permitir-lhe um nível de normalização mais detalhado,

³⁷ A norma Z39.50 foi publicada pela primeira vez em 1988 e, dois anos mais tarde, criava-se o *Z39.50 Implementation Group* (ZIG) e a *Maintenance Agency*, sediada na Biblioteca do Congresso. Estas duas entidades foram responsáveis pela 2ª versão da norma, publicada em 1992 e que, além de incluir aspectos que permitiam uma melhor funcionalidade que tornava a sua implementação economicamente justificada, se aproximava e compatibilizava com o protocolo "Search and Retrieval" (SR), desenvolvido pela ISO entre 1984 e 1991. A 3ª versão do Z39.50 data de 1995 e a sua revisão foi publicada em 2003; a próxima revisão ocorrerá em 2007. Para informação geral sobre a ISO 23950, ver: *International Standard Z39.50 Maintenance Agency* em <http://lcweb.loc.gov/z3950/agency/> e *NISO Z39.50 Resource Page* em <http://www.niso.org/z39.550.html>.

³⁸ O *Bath Profile* : an international Z39.50 specification for library applications and resource discovery especifica um sub-conjunto da norma ANSI/NISO Z39.50 – 1995 (ISO 23950). A

tornando-o mais apto a responder a um maior leque de situações, mesmo às que se colocam no modelo dos catálogos físicos, como no caso de um mesmo catálogo pertencer a dois, ou mais, catálogos colectivos que requeiram diferentes configurações. Em síntese, o *Bath Profile* procurou responder à necessidade de “um perfil relativamente simples no qual os valores de todos os tipos de atributos fossem rigorosamente regulados, de forma a que os sistemas pudessem comportar-se de um modo exacto em resposta a um pequeno número de tipos de questões definidos” (Miller, 1999). As suas especificações permitem a pesquisa inter-domínios numa rede de recursos que agrupa catálogos de biblioteca, informação governamental, sistemas de museus e arquivos. A necessidade de perfis justifica-se pelo facto de a norma (tal como acontece com outras) ter opções de implementação e permitir a interpretação de certas especificações. Assim, quem desenvolve programas pode escolher diversas opções para implantar ou interpretar a norma nos seus sistemas. Como resultado, o utilizador pode receber respostas erradas ou pode mesmo não conseguir extrair uma referência que exista na base de dados. Para que se possa utilizar eficazmente a Z39.50, a versão 1.1 do perfil *Bath* definiu um conjunto de especificações básicas sobre autor, título e assunto de pesquisa e extracção e a versão 2.2 incluiu ainda especificações destinadas à extracção de dados dos fundos documentais a partir de catálogos públicos em linha ou de catálogos colectivos e à pesquisa e extracção de registos de autoridade (Lunau, 2003, p. 2).

Segundo Inês Cordeiro (2001, p. 6), embora algumas posições críticas apontassem a inadequação do Z39.50 ao ambiente tecnológico da *Web*³⁹ e se reconhecesse a sua complexidade em termos de implementação, o protocolo funcionava e deveria continuar a ser explorado enquanto não existissem alternativas equivalentes, salientando-se também os esforços que vinham a ser desenvolvidos no sentido de “alinhamento com a tecnologia *Web*, por exemplo a codificação das mensagens Z39.50 em XML, utilização com metadados

responsabilidade pela sua gestão e desenvolvimento cabe à Biblioteca e Arquivos do Canadá (LAC).

³⁹ O protocolo Z39.50, embora compatível com a Internet uma vez que se apoiava no seu protocolo de base TCP/IP, não o era com a *Web* e o seu protocolo HTTP (Cf. Lahary, 2002, p. 188).

Dublin Core, etc.”⁴⁰. Mais recentemente têm vindo a desenvolver-se iniciativas que apontam para uma maior aproximação à tecnologia *web*, destacando-se o SRW – *Search/Retrieve Web Service*⁴¹, que aproveita os desenvolvimentos dos formatos XML para melhorar, do ponto de vista tecnológico, a possibilidade de realização dos serviços do Z39.50, não pondo em causa a necessidade de manter o protocolo, nem resolvendo todas as limitações que existem na sua utilização, já que “os reais impedimentos para uma cabal obtenção dos resultados esperados com o Z39.50 estão para lá da tecnologia do protocolo e dizem sobretudo respeito a correspondências semânticas e funcionais (como as estruturas de dados, índices, etc. disponíveis em cada sistema) que devem existir, para que o protocolo funcione razoavelmente numa comunidade de sistemas” (Cordeiro e Carvalho, 2003, p. 45-46). Em relação ao tratamento dos assuntos, o perfil *Zthes*⁴² descreve um modelo abstracto para representar e pesquisar tesouros e especifica o modo como o modelo pode ser implementado usando, em especial, os protocolos Z39.50 e SRW⁴³, embora seja igualmente possível implementá-lo com outros protocolos e formatos. E em relação à desejável interoperabilidade entre bibliotecas e museus, parece-nos importante referir o facto de se ter desenvolvido um perfil de aplicação Z39.50 para a informação em património cultural (o perfil CIMI⁴⁴)⁴⁵. Por enquanto, parece ainda válida a afirmação de Carrol Lunau⁴⁶ (2003, p. 1) de que a norma Z39.50 “continua a ser a norma mais correntemente utilizada para realizar pesquisas nos catálogos das bibliotecas e, conseqüentemente, tem um papel a desempenhar na metapesquisa em evolução ou no ambiente de pesquisa de protocolos múltiplos”.

⁴⁰ A adaptação total da norma com o ambiente *Web* e XML é preconizada pelo ZING (Z39.50 International New Generation). Veja-se em especial a informação que consta do seguinte endereço: <<http://www.loc.gov/z3950/agency/zing/zing.html>>.

⁴¹ SRW – *Search/Retrieve web Service* [Em linha]. [Consult. 31 Jan. 2005]. Disponível na WWW: <URL: <http://www.loc.gov/z3950/agency/zing/srw/background.html>>

⁴² As primeiras versões do perfil *Zthes* do Z39.50 datam de 1998 (Cf. Taylor, 2003).

⁴³ *Zthes : a profile for thesaurus navigation in Z39.50 and SRW*. 17th December 2004. [Em linha]. [Consult. 31 Jan. 2005]. Disponível na WWW: <URL: <http://zthes.z3950.org/>>.

⁴⁴ CIMI – *Consortium for the Computer Interchange of Museum Information*.

⁴⁵ *The CIMI profile: Z39.50 application profile for cultural heritage information : Release 1.0* (March 1998). [Em linha]. [Consult. 9 Feb. 2005]. Disponível na WWW: <URL: http://www.cimi.org/old_site/downloads/ProfileFinalMar98/cimiprofile1.htm/>.

⁴⁶ Representante da Biblioteca e Arquivos Nacionais do Canadá (entidade responsável pela manutenção do *Bath Profile*) no primeiro encontro de Bath.

3. A PORBASE – Base Nacional de Dados Bibliográficos: breve caracterização de um catálogo colectivo

3.1 A criação da PORBASE e algumas etapas do seu desenvolvimento como projecto cooperativo

Decorrente do processo de informatização da Biblioteca Nacional iniciado em 1985, a PORBASE - Base Nacional de Dados Bibliográficos, que ficou disponível ao público em Maio de 1988, agrupou bibliotecas interessadas em colaborar no catálogo colectivo em linha das bibliotecas portuguesas, numa resposta à dispersão de informação e iniciando um processo de gradual melhoria no aproveitamento de recursos humanos e materiais, nomeadamente através da partilha de recursos bibliográficos e do desenvolvimento de políticas de aquisições partilhadas e do empréstimo inter-bibliotecas. Tratando-se de uma base aberta “à cooperação espontânea de cada biblioteca” (Cabral, 1987, p.164), a Base Nacional de Dados Bibliográficos – PORBASE veio reforçar o papel da Biblioteca Nacional enquanto Agência Bibliográfica Nacional, na sua faceta eminentemente normalizadora. O desenvolvimento de todo um conjunto de iniciativas no âmbito da normalização – de que se destacam, a nível da descrição bibliográfica, a implementação das *Regras Portuguesas de Catalogação*, editadas em 1984 pelo Departamento de Bibliotecas, Arquivos e Serviços de Documentação do então Instituto Português do Património Cultural, a adopção do formato UNIMARC como formato de descrição nacional⁴⁷ e a elaboração de regras para as entradas de autoridade e a sua correspondência ao formato Unimarc/Autoridades, cujo manual, traduzido para português, foi editado em 1989⁴⁸, após uma edição prévia em folhas móveis.

Mas não foi apenas a nível da descrição física que o papel normalizador da Biblioteca Nacional, no contexto da Base Nacional de Dados Bibliográficos, se fez sentir. Também a nível do tratamento dos conteúdos se desenvolveram

⁴⁷ O *Manual Unimarc* foi traduzido para língua portuguesa em 1989, depois de uma edição em folhas volantes destinada a apoiar a catalogação feita a partir de folhas de recolha de dados (FRDs).

⁴⁸ Esta edição corresponde a uma versão provisória, pois o formato encontrava-se então ainda em discussão no seio do *IFLA Steering Group on Unimarc Format for Authorities*. A versão final, em língua inglesa, só seria editada em 1991.

sistemas e instrumentos de trabalho que permitiram uma melhor representação dos conteúdos temáticos da PORBASE. Enquadram-se neste esforço a edição, em 1987, de uma *Tabela de autoridade CDU*, contendo uma selecção das tabelas gerais daquele sistema de classificação e das notações de cada classe então consideradas como suficientes para um nível mínimo de descrição exigida às bibliotecas cooperantes, já que a Classificação Decimal Universal (CDU) foi adoptada como linguagem documental comum para a representação de conteúdos nos registos bibliográficos integrados na PORBASE. Paralelamente, foi desenvolvido por um grupo de trabalho da Biblioteca Nacional o SIPORbase – Sistema de Indexação em Português, um conjunto de formulações teórico-práticas, oferecendo um corpo de regras e instruções para a selecção de terminologia e sua utilização de acordo com os princípios da pré-coordenação. Abria-se assim o caminho a uma prática até aí inexistente na Biblioteca Nacional, a da indexação terminológica. Embora este sistema possa ser uma opção no tratamento de assuntos por parte das bibliotecas portuguesas, na realidade a maioria que faz indexação continua a regular-se por linguagens que tendem para a pós-coordenação.

As diferentes opções a nível das linguagens de indexação adoptadas pelas bibliotecas cooperantes da PORBASE, assim como a variedade de instrumentos de indexação utilizados, provocavam a inconsistência do índice de assuntos e demonstraram a necessidade de normalização da Base Nacional de Dados Bibliográficos a nível do acesso por assunto. Perante esta situação e reconhecidos os inconvenientes que ela causava ao utilizador, foi suspensa a integração dos campos de indexação dos registos bibliográficos das bibliotecas cooperantes, aceitando-se apenas para carregamento o campo da CDU (675 do formato Unimarc). Um primeiro passo para tentar resolver o problema da pesquisa por assunto na PORBASE foi a já referida adopção da Classificação Decimal Universal (CDU) como linguagem documental comum e, como tal, a obrigatoriedade da sua utilização pela comunidade dos cooperantes. No entanto, conhecidas as limitações que a CDU tem a nível da representação de assuntos, lançou-se, em 1989, na sequência das 1^{as} Jornadas PORBASE, um projecto com o objectivo de normalizar e compatibilizar a terminologia de indexação usada para representar os assuntos nas diferentes áreas temáticas

– trata-se do **Projecto CLIP** – Compatibilização de Linguagens de Indexação em Português. Segundo o seu *Regulamento*, este "é um projecto cooperativo de concepção e gestão integrada de termos de indexação para aplicação pelas bibliotecas portuguesas participantes na PORBASE – Base Nacional de Dados Bibliográficos"⁴⁹. Estruturado em dezasseis grupos de trabalho, a cada um dos quais foi acometida uma área temática e que integram representantes de bibliotecas especializadas (universitárias ou não) e, em alguns casos, outros especialistas, para além de representantes da Biblioteca Nacional, foi sua tarefa inicial o levantamento das fontes existentes para o controlo da terminologia em cada um dos domínios temáticos⁵⁰, face ao que a metodologia adoptada por cada grupo foi bastante variável. Se em áreas como a Agricultura e as Ciências da Saúde existiam já instrumentos de indexação largamente utilizados nas bibliotecas especializadas, quer a nível nacional quer a nível internacional, como o *Agrovoc* no primeiro caso e os *Medical Subject Headings* para o segundo (na versão em língua portuguesa traduzida pelo BIREME - Brasil), adoptaram-se essas linguagens como válidas no âmbito do Projecto CLIP e, como tal, passíveis de integração as PORBASE; já noutros casos foi necessário recorrer às várias fontes identificadas e desenvolver, com base nelas ou não, mas sempre tomando-as em consideração, todo um trabalho de criação das terminologias para indexação das áreas específicas. Estão neste caso, por exemplo, as áreas do Ambiente e das Belas Artes. Para lá do seu trabalho de adopção/adaptação/criação de uma terminologia para indexação, alguns dos grupos desenvolveram, entre meados de 1994 e meados de 1996, um trabalho de validação de registos de autoridade/assunto já existentes na PORBASE, tendo incidido essa validação sobre cerca de 2 000 registos (Santos, Luísa, 2000). A actividade destes grupos e a consequente obtenção de resultados não foi idêntica para todos. Se alguns produziram terminologias e validaram termos, outros não chegaram sequer a funcionar, tornando-se necessário reformular e relançar alguns desses grupos, como foi, aliás, reconhecido nas Jornadas Porbase realizadas em Março de 2000.

⁴⁹ *Regulamento do Projecto CLIP*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.porbase.org/coop/temp.asp?id=778&idm=849> >

⁵⁰ Este trabalho resultou na publicação de *Fontes para o controlo de terminologia*. Lisboa: Biblioteca Nacional, 1995. Este inclui, além das fontes, a composição dos grupos de trabalho assim como a metodologia adoptada por cada um deles.

Paralelamente a todo este esforço normativo, a Biblioteca Nacional assumiu também o papel de Centro Distribuidor Nacional do CDS/ISIS, apostando desde cedo no desenvolvimento e distribuição de ferramentas informáticas que facilitassem a modernização das bibliotecas portuguesas e permitissem a sua mais rápida integração na comunidade das bibliotecas cooperantes da PORBASE. As várias parametrizações Porbase do Mini-Micro CDS/ISIS (quatro, algumas das quais com mais de uma versão), concebidas para MS-DOS, foram produzidas e distribuídas até 2000, sendo então substituídas pelo novo sistema Porbase 5, concebido para funcionar em ambiente Windows.

O desenvolvimento da PORBASE, desde 1987, passou por vários momentos significativos, de que podemos destacar o enquadramento que lhe foi dado pela lei orgânica que criou o Instituto da Biblioteca Nacional e do Livro a partir da fusão da Biblioteca Nacional e do Instituto Português do Livro e da Leitura, em 1992⁵¹, a disponibilização da base de dados na Internet em 1995⁵², e a alteração do sistema informático de suporte da PORBASE, o GEAC 9000, assente no sistema informático GLIS (*Geac Library Information System*) pelo sistema Horizon, em 1998. Este sistema, que representa os princípios das bases de dados relacionais (sistema Sybase), é aberto, modular, integrado, com processamento em tempo real e baseia-se numa filosofia cliente/servidor, utilizando o nível 4 da OSI⁵³, multi-sistema e multiplataforma que corre em SQL⁵⁴ *Server*. Dos diferentes módulos que o compõem, destacam-se o de

⁵¹ Ao criar a Direcção de Serviços da PORBASE, permitiu um desenvolvimento mais cabal de actividades e projectos inerentes à gestão e desenvolvimento da Base Nacional de Dados Bibliográficos (ver Decreto-Lei n.º106-E/92, de 1 de Junho). Pelo Decreto-Lei n.º89/97, de 19 de Abril, a Biblioteca Nacional recuperou a sua autonomia e a coordenação das acções de desenvolvimento da PORBASE enquanto catálogo colectivo foi acometida à Divisão da PORBASE, integrada na nova Direcção de Serviços de Inovação e Desenvolvimento.

⁵² Em 1994-95 o sistema GEAC 9000 foi dotado de "facilidades tecnológicas para o tornarem um sistema aberto a outras redes de computadores" (Lopes, 1995), o que, aliado à cooperação com entidades como a Fundação para o Desenvolvimento dos Meios Nacionais de Cálculo Científico – FCCN, contribuiu para a disponibilização da base de dados na Internet.

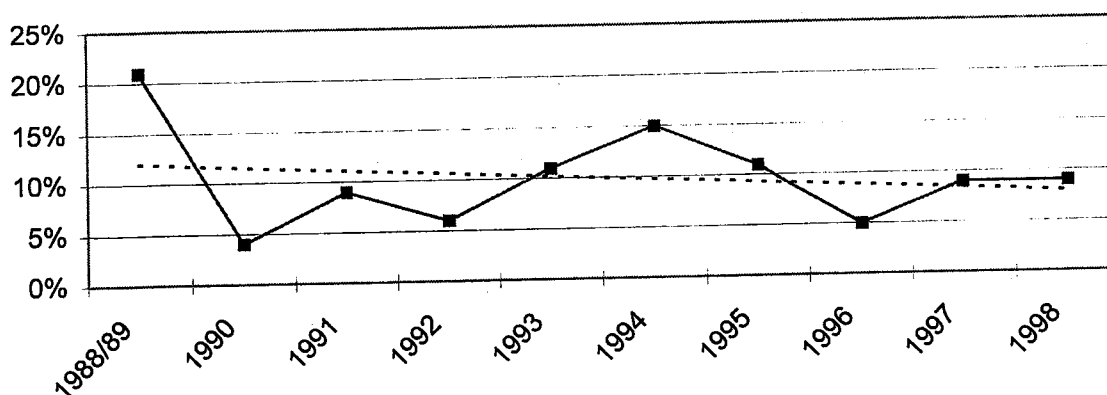
⁵³ Open Systems Interconexion (ISO/IEC 7498-1.1994, Information technology – *Open Systems Interconexion – Basic Reference Model: the Basic Model*).

⁵⁴ *Structured Query Language* - trata-se de uma linguagem de interrogação normalizada. Para mais detalhes técnicos referentes ao sistema Horizon e a sua implementação na Biblioteca Nacional, veja-se BIBLIOTECA NACIONAL. Grupo de Implementação do Sistema Horizon - PORBASE : Novas perspectivas de desenvolvimento. JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000]

Aquisições, o de Catalogação e Controlo de Autoridade, o de Gestão de Periódicos, o módulo de Pesquisa Profissional, o OPAC e o módulo Internet (inicialmente Webpac), mais tarde reunidos no iPAC, substituído por seu turno em Janeiro de 2005 pelo HIP (*Horizon Information Portal*) e, ainda, o módulo de Circulação e Empréstimo e o ILL (*Inter Library Loan*). Ao implementar as normas e recomendações sobre a interconexão de sistemas abertos, nomeadamente os protocolos Z39.50 e ILL, o novo sistema abria novas possibilidades no acesso a outras bases de dados e abria o caminho a outras formas de cooperação, não exclusivamente baseadas, como até aí, no carregamento em diferido ou, nalguns casos (poucos), em linha, de registos oriundos das bibliotecas cooperantes.

Até 1998 (durante os dez anos em que o sistema GEAC suportou a PORBASE), o ritmo de integração dos registos, incluindo os da Biblioteca Nacional, referentes a um universo de 1 033 227 registos bibliográficos (n.º de registos existente na base no final de 1998), foi o que consta do seguinte quadro:

Integração de registos na PORBASE (1988-1998) - em percentagem

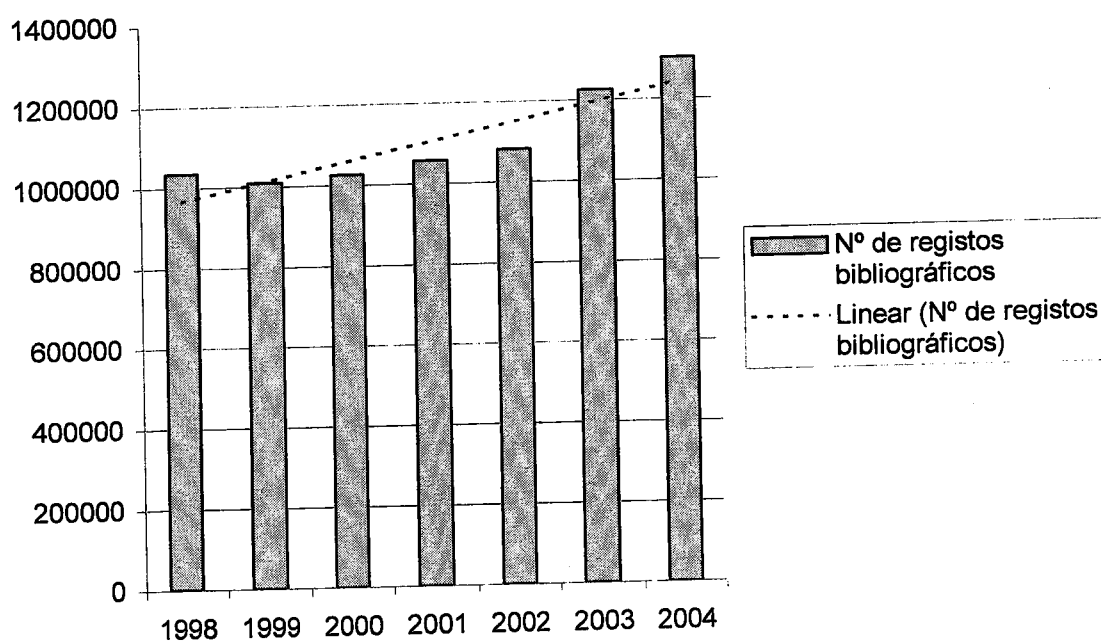


Fonte: GALVÃO, Rosa Maria - Desenvolvimento da base de dados : apresentação das actividades e resultados da cooperação. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000].

Com a entrada em funcionamento do novo sistema Horizon assistiu-se a uma “diminuição” do número de registos bibliográficos, resultante da reorganização dos registos existentes na base; com efeito, o número de controlo bibliográfico

(ncb) daqueles que, durante a vigência do GEAC, tinham sido eliminados, continuava a existir e só desapareceu com a integração no novo sistema (Galvão, 2000). Ainda como resultado de todo o trabalho de implementação do sistema Horizon, foi interrompida, a partir do 2º trimestre de 1999, a integração em diferido dos registos das bibliotecas cooperantes, integração essa que só foi efectivamente retomada no 1º trimestre de 2003 (e que o gráfico seguinte atesta):

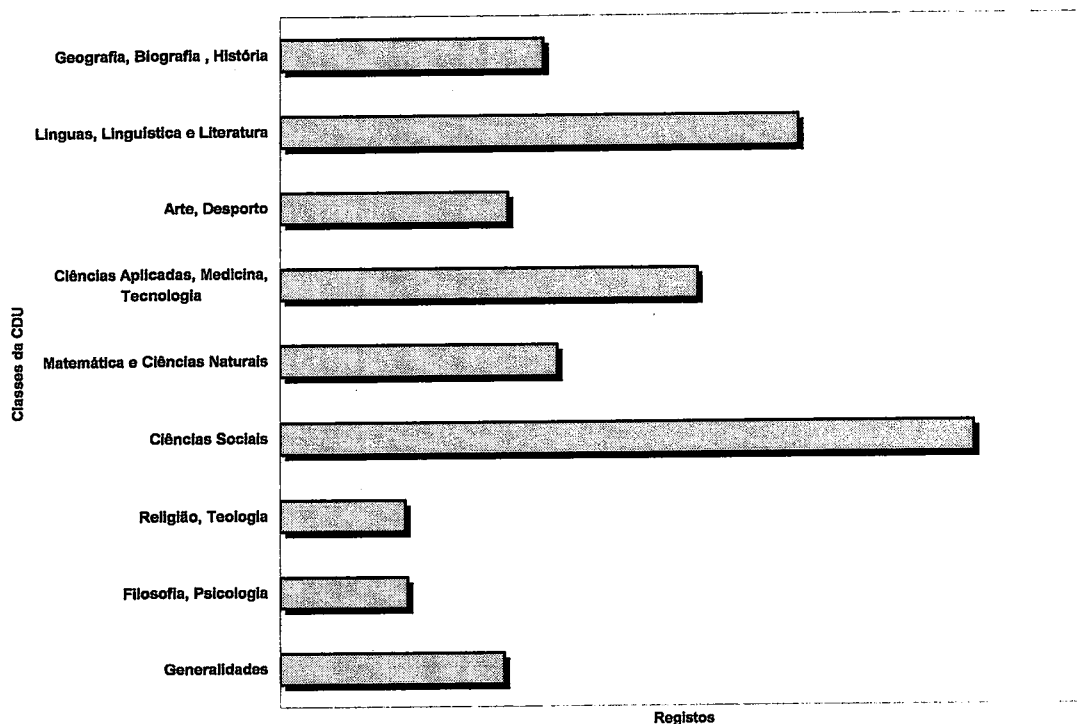
Evolução da base de dados 1998-2004



Fonte: <URL: http://stats.porbase.org/Default_BN.asp?lang=por>[Consult. 5 Mar. 2005].

3.2 Caracterização do conteúdo da base de dados e do universo das bibliotecas cooperantes

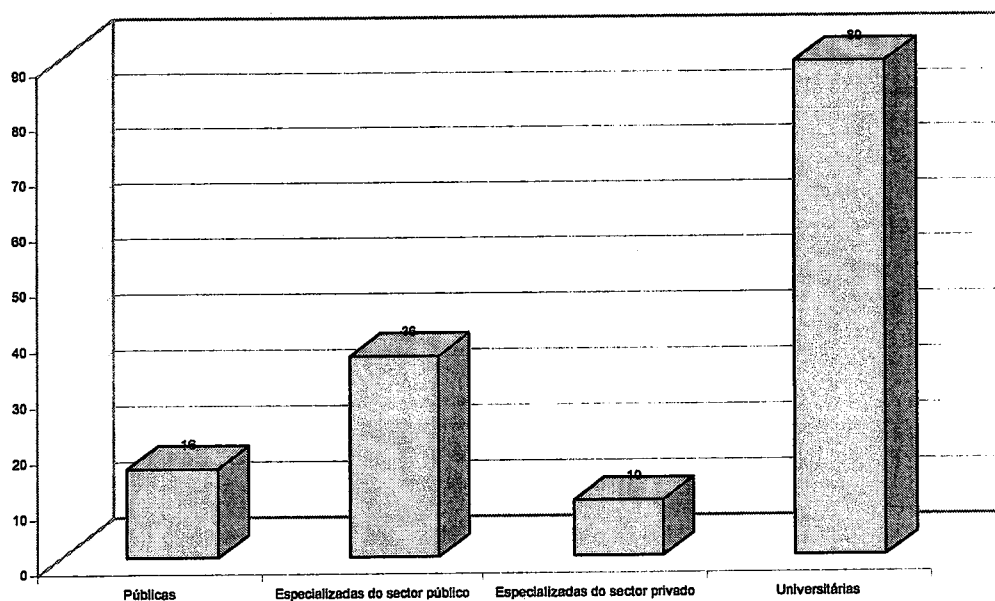
Em relação aos conteúdos temáticos, é a seguinte a composição da base de dados:



Fonte: BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório de Actividades de Cooperação 2000/2002*. Lisboa: BN, 2003.
 [Em linha] <<http://www.porbase.org/relatorio/relatorio2000-2002.doc>>

O universo das bibliotecas cooperantes, cujo principal grupo é constituído pelas bibliotecas universitárias, é o seguinte:

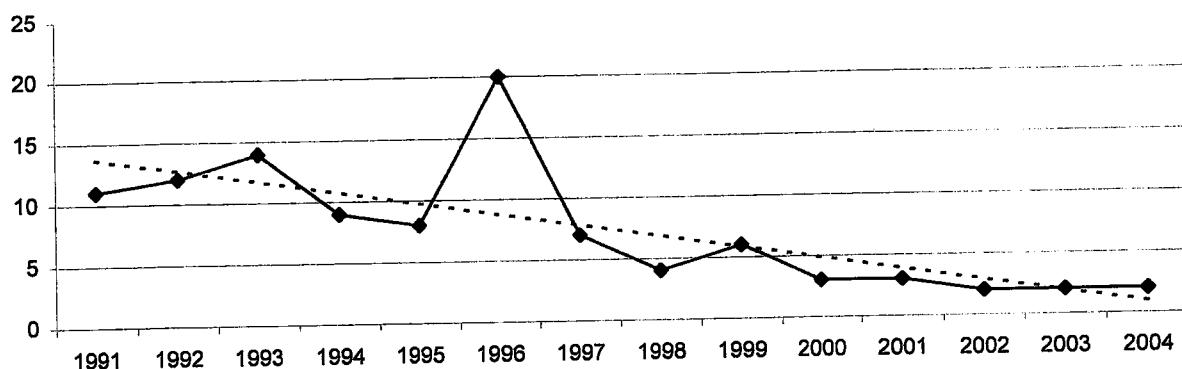
Tipos de Biblioteca Cooperante (excepto a Biblioteca Nacional)



Fonte: BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório de Actividades de Cooperação 2000/2002*. Lisboa: BN, 2003.
 [Em linha] <<http://www.porbase.org/relatorio/relatorio2000-2002.doc>>

O ritmo de adesões à PORBASE – Base Nacional de Dados Bibliográficos entre 1988 e 2004 pode ilustrar-se com o seguinte gráfico:

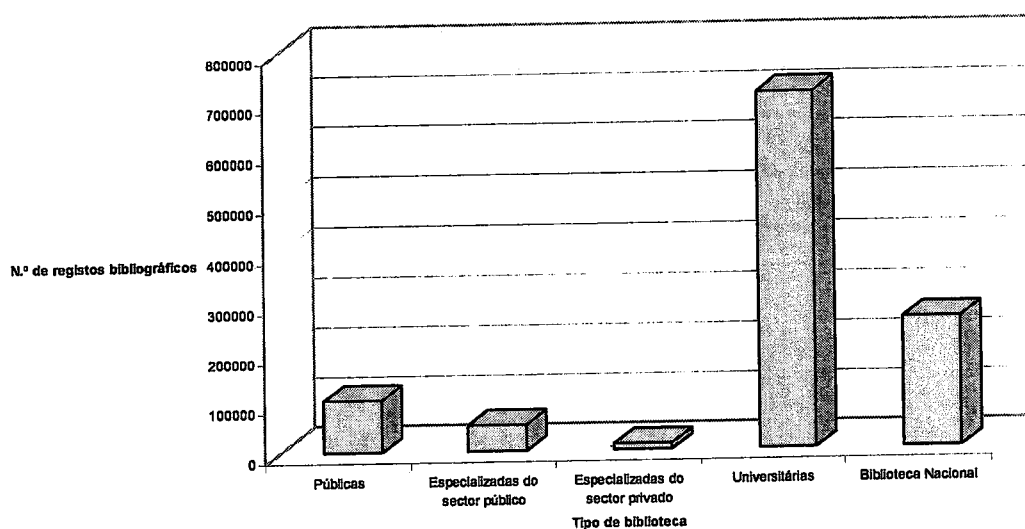
Evolução das adesões à PORBASE - Base Nacional de Dados Bibliográficos



Fontes: SANTOS, Lúcia - A contribuição das bibliotecas cooperantes para o crescimento da PORBASE – Base Nacional de Dados Bibliográficos. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000]. BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório das actividades de cooperação 2000-2002*. [Em linha] <http://www.porbase.org/relatorio/relatorio2000-2002.doc>. BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório de actividades 2003*. BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório de actividades 2004*.

Quanto à contribuição de registos para a Base Nacional de Dados Bibliográficos, podemos sintetizá-la no seguinte quadro:

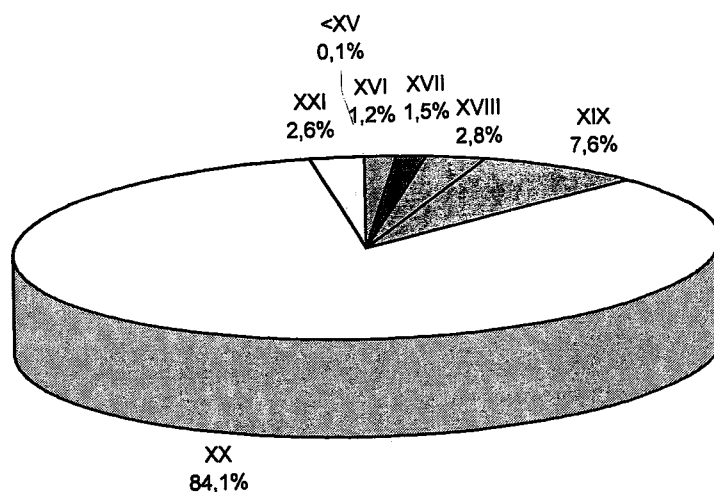
Contribuições para a PORBASE por tipo de biblioteca



Fonte: BIBLIOTECA NACIONAL. Divisão da PORBASE – *Relatório das actividades de cooperação 2000-2002*. [Em linha] <<http://www.porbase.org/relatorio/relatorio2000-2002.doc>>

Em relação ao conteúdo da base de dados no que se refere à data de edição dos documentos nela referenciados, a situação é a seguinte:

Conteúdos da PORBASE por data de edição (séculos)



Fonte: <URL: http://stats.porbase.org/Default_BN.asp?lang=por [Consult. 5 Mar. 2005].

Este gráfico revela-nos que ca 5,5% dos registos bibliográficos existentes correspondem a Livro Antigo, o que constitui um núcleo a desenvolver na óptica da criação de uma base de dados específica para este tipo de material, independentemente da sua referenciação na PORBASE.

Após dezassete anos de funcionamento, podemos dizer que a PORBASE - Base Nacional de Dados Bibliográficos é constituída por mais de um milhão e trezentos mil registos bibliográficos, vindos maioritariamente do grupo das bibliotecas universitárias, e que no seu conteúdo temático se destaca a área das Ciências Sociais, como vimos nos gráficos anteriores. É igualmente de assinalar a tendência decrescente do n.º de adesões à PORBASE, assim como os "picos" de 1988-89, 1992-93 e de 1996. Estes podem compreender-se, pelo menos em parte, pelo próprio contexto de evolução da PORBASE: em 1988-89 iniciava-se a cooperação e as vantagens evidentes que a PORBASE apresentava a um universo de bibliotecas que iniciavam o seu processo de informatização levaram a que o nível de adesão fosse bastante assinalável; em 1993 estruturava-se, na Biblioteca Nacional, a recém-criada Direcção de

Serviços da PORBASE; e, em 1996, eram evidentes as vantagens que a disponibilização da Base Nacional de Dados Bibliográficos na Internet, através referida, oferecia ao universo de bibliotecas que nela participassem.

No que respeita às áreas temáticas, a PORBASE reflecte os conteúdos das bibliotecas que mais registos enviam para integração, com larga predominância das Ciências Sociais, como vimos, da Literatura e da Linguística. Esta realidade pode estar relacionada com as características da actividade editorial portuguesa respeitante às publicações impressas (que se reflecte nas obras entradas na Biblioteca Nacional por Depósito Legal), questão que não aprofundámos mas que poderá constar de uma abordagem às relações entre o mundo da edição, as bibliotecas e os leitores, *latu sensu*.

Por outro lado, a realidade espelhada na PORBASE não considera um tipo de documento extremamente importante para a investigação e o conhecimento em geral - os analíticos, sejam eles de monografias (como as comunicações que constam das actas de congressos e outras reuniões científicas, por exemplo), ou de publicações em série. Pela vastidão temática que cobrem, só muito dificilmente a Biblioteca Nacional poderia possuir os recursos necessários a esse nível de tratamento das publicações que dão entrada por Depósito Legal. Desde o início da PORBASE que se tem tentado ultrapassar esta situação, procurando desenvolver projectos de catalogação partilhada de analíticos que pudessem vir a ser integrados na base, mas as alterações entretanto ocorridas a nível do bloco 4XX do formato Unimarc / Bibliográfico (que diz respeito às entradas relacionadas) vieram dificultar a sua concretização, pois obrigaram as bibliotecas e os próprios sistemas de gestão das bases de dados bibliográficos a adaptarem-se às novas condições. Neste momento, estão criadas as condições técnicas para se avançar com um projecto que possibilite tratar e disponibilizar esse importante acervo informativo.

3.3 A PORBASE e os modelos de catálogo colectivo

Retomando a questão dos modelos de catálogos colectivos, parece-nos evidente que a PORBASE – Base Nacional de Dados Bibliográficos possui uma

arquitectura de catálogo colectivo centralizado/concreto, em que “as bibliotecas do grupo partilham o acesso a um único sistema automatizado e a uma única base de dados que é, simultaneamente, o catálogo de todas elas e o catálogo colectivo de todas” (Campos, 1995, p.5), independentemente destas bibliotecas poderem ou não ter os seus próprios catálogos e sistemas. Prosseguindo na linha de análise de Dovey (2000), parece-nos que a PORBASE, de certa forma, preenche os requisitos dos seus quatro tipos de modelos físicos: para a Biblioteca Nacional, o catálogo colectivo funciona igualmente como seu catálogo principal; os registos das bibliotecas cooperantes, na sua maioria, são enviados para o Órgão Coordenador da PORBASE, que procede ao seu carregamento em diferido; a produção da Bibliografia Nacional Portuguesa, editada em CD-ROM entre 1995 e 2002 e, a partir de 2003, acessível em linha, permite às bibliotecas (sejam ou não cooperantes) retirarem daí registos para as suas bases de dados locais, o que é, desde o início de 2005, igualmente possível directamente a partir da PORBASE, e cumpre as condições deste tipo de catálogo físico definido por Dovey; e, finalmente, a actualização simultânea do catálogo central e de catálogos locais tem sido de certo modo conseguida em relação às bibliotecas que catalogam em linha, num conceito de catalogação distribuída, passível de vir a ser alargada aproveitando o facto de o sistema Horizon implementar o protocolo Z39.50.

Até ao presente, têm sido estas as características da Base Nacional de Dados Bibliográficos; contudo, desde 1995 que se lhe aponta o caminho do catálogo virtual (Campos, 1995; Lopes, 1995) como aquele que melhor se adapta às novas exigências da difusão da informação e da partilha de recursos e que, ao exigir soluções tecnológicas compatíveis com esta nova realidade, ao alterar determinados procedimentos técnicos e ao propor novos processos de disponibilização e acesso à informação, poderá igualmente propiciar uma mudança da filosofia que, até agora, tem estado subjacente ao desenvolvimento da PORBASE, implementando de forma mais profunda uma vertente de serviço orientado para grupos de utilizadores, sem deixar, contudo, de crescer como um importante repositório de informação existente nas bibliotecas portuguesas.

4. O acesso por assunto à informação

4.1 Os modelos tradicionais de representação do conhecimento e de recuperação de assuntos

Na tentativa de organizar o conhecimento materializado nos fundos documentais das bibliotecas, desenvolveram-se diferentes sistemas com o objectivo de lhes fornecer uma estrutura que facilitasse a disposição e arrumação física dos documentos e permitisse igualmente representar os assuntos neles contidos em expressões formais que pudessem ser fácil e rapidamente apreendidas e interpretadas na recuperação. Os vários tipos de classificações (universais, gerais ou especializadas, quanto à abrangência temática, ou enumerativas, hierárquicas, de facetas, quanto à organização e estrutura) e as diversas abordagens terminológicas (consubstanciadas em tesouros de carácter geral ou, mais frequentemente, especializados, em listas de cabeçalhos de assunto e noutros instrumentos de apoio à indexação), embora reflectam diferentes fundamentos epistemológicos e diferentes perspectivas teóricas e metodológicas da análise, organização e representação dos assuntos, têm coexistido no mundo biblioteconómico e os seus desenvolvimentos têm, por vezes, revelado algumas influências mútuas. Neste capítulo iremos abordar, de forma sucinta, apenas aqueles sistemas mais genericamente utilizados pelas bibliotecas, procurando definir as suas características específicas e a forma como estas se adaptam, num grau maior ou menor de eficácia, às necessidades da recuperação da informação.

4.1.1 As classificações universais

As classificações são os mais antigos instrumentos de organização do conhecimento usados pelas bibliotecas. Se as primeiras tentativas de organizar os seus acervos remontam ao séc. XVI⁵⁵, é certo que só no século XIX, com o desenvolvimento da Ciência e a sua organização de acordo com sistemas

⁵⁵ Conrad Gesner (1516-1565), com a sua *Bibliotheca Universalis* escrita entre 1545 e 1555, além de referenciar e resumir cerca de 16 000 títulos de obras, manuscritas e impressas, sobre todas as áreas do conhecimento e escritas em latim, grego e hebraico, apresentava também (em 1548) uma versão metódica, com as referências classificadas pelos assuntos (Cf. *Tous les savoirs du monde : liste des textes fondateurs : quelques encyclopédies et textes fondateurs de la pensée occidentale, avant l'explosion des savoirs*. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://classes.bnf.fr/dossitsm/textfond.htm#XVIe>>.

estruturados, por uma lado, e as novas possibilidades oferecidas pela tipografia industrial que, ao permitir o desenvolvimento da imprensa e da actividade editorial conduziu igualmente ao crescimento das colecções das bibliotecas, por outro, surgem as grandes classificações universais, ainda hoje importantes instrumentos de organização e de representação dos conteúdos temáticos dos documentos existentes e/ou disponibilizados pelas bibliotecas. Estas classificações, de que destacamos a Classificação Decimal de Dewey (CDD), publicada pela primeira vez em 1876, e a Classificação Decimal Universal (CDU), cuja primeira edição data de 1904-1907⁵⁶, adequavam-se não só à representação, numa linguagem uniforme, dos principais assuntos dos documentos, mas também à sua organização física nas estantes e salas das bibliotecas, de acordo com as várias temáticas. São ambas classificações de tipo hierárquico, que têm subjacente uma organização sistemática do conhecimento cuja estrutura reflecte a época em que foram concebidas. Contudo, ao longo do tempo elas foram objecto de adaptações sucessivas às novas realidades científicas e acompanharam a evolução tecnológica e as novas exigências das bibliotecas, transformando as suas versões impressas, de morosa actualização, em instrumentos em linha facilmente acedíveis e actualizáveis, conseguindo assim não só manter mas também expandir a sua utilidade enquanto sistemas estruturados de organização do conhecimento⁵⁷.

A estrutura decimal destas classificações permite-lhes desenvolver cada notação (representativa de um assunto) até elevados níveis de especificidade, ainda que o resultado possa ser uma cadeia de números e símbolos dificilmente compreensíveis pelos utilizadores. Apesar de ambas possuírem esta estrutura, há algumas diferenças fundamentais que as distinguem, entre as quais o facto de a CDD ter mantido uma característica marcadamente enumerativa e a CDU, três vezes maior que a CDD, ter evoluído como uma classificação sintética (Slavic, 2003, p.3), com evidentes vantagens a nível da

⁵⁶ Cf. <URL: <http://www.udcc.org/about.htm>>.

⁵⁷ As actualizações da Classificação Decimal de Dewey são hoje responsabilidade da OCLC Forest Press (<URL: <http://www.oclc.org/dewey/>>), apoiada por um comité editorial de 10 membros, representantes de bibliotecas de todo o mundo, entre as quais a *British Library*, a *Library of Congress* e a *Library and Archives of Canada*. A Classificação Decimal Universal é mantida por um consórcio internacional, o *UDC Consortium* (<URL: <http://www.udcc.org/>>), composto por organismos de normalização, bibliotecas e entidades editoras, igualmente apoiado por um conselho consultor de especialistas internacionais.

cobertura de assuntos (em número ilimitado) e da sua combinação. Esta característica confere-lhe também uma maior complexidade, já que requer o conhecimento de regras sintáticas e também um maior suporte a nível das ferramentas de gestão da linguagem (Slavic, 2003, p.7). A cadeia notacional que construímos quando classificamos um assunto (com base nas tabelas principais, para as classes e subclasses e nas tabelas auxiliares, para os auxiliares comuns) corresponde, na realidade, a um termo de indexação pré-coordenado, em que cada elemento tem o mesmo significado quer se encontre isolado, quer conste de uma qualquer combinação. Por exemplo, na notação 72(469)“19”(043.2), que representa uma tese (043.2) sobre arquitectura (72) em Portugal (469) no século XX (“19”), cada elemento tem um significado único, que naquela combinação contribui para especificar um determinado assunto, mas que numa combinação diferente ajudará a representar um assunto também muito diferente. Esta característica torna a CDU uma linguagem adequada à pesquisa pós-coordenada (em que cada elemento poderia ser recuperado individualmente ou nas suas mais diversas combinações). Contudo, nos esforços que têm vindo a ser desenvolvidos para que as classificações possam ser cada vez mais manipuláveis e reutilizáveis por programas informáticos e apesar dos avanços que o *Master Reference File* da CDU (UDC MRF)⁵⁸ representa, não existe ainda um formato que permita realizar aquele tipo de pesquisa, já que o formato MARC 21 para classificação desenvolvido pela Biblioteca do Congresso está concebido tendo como objectivo a gestão da sua classificação e da própria CDD (ambas enumerativas) e, como tal, não responde a todas as necessidades da CDU. Por outro lado, o formato UNIMARC/Classificação, desenvolvido por iniciativa do Comité Permanente do UNIMARC e cuja versão provisória, ainda incompleta, data de 2000⁵⁹, não avançou muito mais do que o formato MARC 21 a nível da adaptação às características pré-coordenadas da CDU, em particular quanto à necessidade de pesquisar individualmente os elementos significativos que compõem uma cadeia notacional e em relação à gestão de mudanças globais que ocorram a nível desses elementos (Slavic, 2003, p. 11).

⁵⁸ O manual do *Master Reference File* (versão de Fevereiro de 2003) está disponível em <URL: <http://www.udcc.org/mrf.htm>>

⁵⁹ Formato abreviado (*Concise UNIMARC Classification Format*), na versão de 31.10.2000, disponível em <URL: <http://www.ifla.org/VI/3/p1996-1/concise.htm>>

Para além destas duas classificações universais, uma classificação geral, a da Biblioteca do Congresso⁶⁰, começou a ser concebida entre 1899 e 1920 com o objectivo de classificar as colecções daquela biblioteca, ainda que posteriormente tenha vindo a ser adoptada por muitas outras. Construída com aquele propósito, ela foi assim reflectindo os conteúdos temáticos do acervo da Biblioteca do Congresso, apresentando por isso um desigual desenvolvimento das áreas temáticas.

Sempre com o objectivo de permitir uma identificação rápida de domínios e subdomínios temáticos, surgiram ainda outras classificações mais especializadas, como, por exemplo, o esquema de categorização AGRIS⁶¹, a classificação da ACM (*Association for Computer Machinery*)⁶² ou a classificação *EI*⁶³.

4.1.2 Os sistemas terminológicos: tesouros e listas de cabeçalhos de assunto

Os tesouros terão surgido pela primeira vez em 1958⁶⁴ e, uma vez que utilizavam uma linguagem controlada mais próxima da linguagem natural, por ser terminológica e não notacional ou simbólica, pareciam mais próximos de poder satisfazer as necessidades crescentes de informação por permitirem um tratamento dos conteúdos temáticos de um modo mais específico, ao nível do conceito, sem que se perdesse a rede de relações semânticas do termo correspondente ao conceito por ele representado. Com efeito, a aceleração da circulação de saberes aliada ao desenvolvimento da investigação científica e aos novos meios de difusão do conhecimento fizeram com que o tesouro se tornasse o instrumento por excelência dos serviços de documentação que assentavam grande parte da sua actividade no tratamento e difusão da

⁶⁰ A estrutura desta classificação está disponível em <URL: <http://www.loc.gov/catdir/cpsol/lcco/lcco.html>>.

⁶¹ Disponível em <URL: <http://www.fao.org/agris/>>.

⁶² A estrutura do *ACM Computing Classification System* está disponível em <URL: <http://www.acm.org/class/1998/ccs98.html>>.

⁶³ *Engineering Information*.

⁶⁴ NEEDHAM, R. M. e JOYCE, T. – *The thesaurus approach to information retrieval*. "American documentation, 9(3), 1958, p. 192-197. Cf. MENON, Bruno – *L'évolution des langages documentaires*. 2004. [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW: <URL: <http://www.bmenon.net/Evolution%20des%20langages%20documentaires.htm>>.

informação científica e técnica, assente em grande parte nas publicações periódicas. O desenvolvimento paralelo da informática, permitindo operações de pesquisa e recuperação como a que utiliza a lógica booleana (com a possibilidade de intersecção, união e complemento), potenciaram a utilidade deste novo instrumento de indexação, assente nos pressupostos da pós-coordenação. A sua organização, fundada num sistema de relações entre termos representativos de conceitos, com controlo de sinonímia e de equivalência entre termos através de referências de substituição, com uma estrutura semântica composta por termos genéricos, termos específicos e termos relacionados, adequava-se a uma utilização ampla por parte dos serviços de informação por facilitar a escolha dos termos de indexação mais pertinentes e por permitir a sua contextualização, a nível da linguagem documental.

Para Elaine Svenonius (2004, p. 584), os tesauros surgiram na convicção de que as complicadas regras sintácticas dos cabeçalhos de assunto e das classificações seriam substituídas pelos operadores booleanos, mas mais tarde, quando as limitações destes se tornaram evidentes, foi necessário introduzir os operadores de proximidade para explorar as relações contextuais. Considerando que os tesauros, por serem limitados em relação ao que podem expressar, são também limitados enquanto representações do conhecimento, Svenonius considera-os como pouco capazes de permitir precisão na recuperação. Confrontando os tesauros com a CDD, considera esta uma linguagem que, “usando uma codificação notacional, pode expressar assuntos complexos melhor do que mesmo os sistemas baseados em palavras”, pode, “em função das suas hierarquias de perspectiva, exprimir uma grande quantidade de informação relacional, mais do que a que pode ser expressa pelos tesauros tradicionais” (2004, p. 584), numa linha que defende a maior adequação das linguagens altamente formalizadas (como são as classificações) aos objectivos da automatização e da indexação distribuída.

Tal como aconteceu com as classificações gerais, também os tesauros procuraram acompanhar o desenvolvimento tecnológico, criando versões

informatizadas e em linha⁶⁵, permitindo uma actualização mais fácil e uma utilização mais ampla.

Além das linguagens terminológicas inseridas em tesouros, que podem ser especializados ou gerais e que se prestam a uma recuperação assente na pós-coordenação, existem também as listas de cabeçalhos de assunto, geralmente ligadas a sistemas de indexação pré-coordenados. A mais conhecida destas linguagens é a LCSH – *Library of Congress Subject Headings*⁶⁶, que inspirou directamente outras do mesmo tipo, como a do *Répertoire de vedettes-matière* (RVM)⁶⁷ da Biblioteca da Universidade de Laval (Québec, Canadá) e a linguagem RAMEAU⁶⁸ dela derivada, além de ter servido também de base conceptual do SIPORbase - Sistema de Indexação em Português⁶⁹. Se as linguagens extraídas dos tesouros permitem uma indexação mais rápida dos itens a tratar, sendo a recuperação dos vários conceitos representados feita *a posteriori*, as linguagens pré-coordenadas, embora implicando menor rapidez no processo de indexação, possibilitam uma maior precisão na descrição dos assuntos complexos e permitem uma percepção mais imediata, por parte do utilizador, do conteúdo temático dos itens. Estas linguagens são adequadas às necessidades de bibliotecas de carácter geral, cujos acervos reflectem uma grande diversidade temática, mas também a bibliotecas especializadas. Podendo atingir níveis de especificidade exactamente equivalentes ao assunto que pretendem descrever, apresentam igualmente, com os sistemas automatizados, as vantagens associadas à pós-coordenação, já que os elementos que constituem os cabeçalhos pré-coordenados podem ser pesquisados individualmente (mantendo sempre a sua função semântica) ou enquanto elementos de qualquer cabeçalho complexo. A inserção destas linguagens em ficheiros de autoridade ligados aos ficheiros bibliográficos

⁶⁵ Como o *Eurovoc thesaurus* (<URL: <http://europa.eu.int/celex/eurovoc/>>) e o *Agrovoc* (<URL: <http://www.fao.org/agrovoc/>>), para citar apenas dois exemplos.

⁶⁶ Informações sobre esta classificação podem obter-se em <URL: <http://www.loc.gov/library/libarch-thesauri.html>>.

⁶⁷ Sobre esta linguagem, veja-se o <URL: <http://www.collectionscanada.ca/rvm/index-e.html>>.

⁶⁸ Esta linguagem pode ser consultada em <URL: <http://rameau.bnf.fr/>>.

⁶⁹ PORTUGAL. Biblioteca Nacional. Área de Classificação e Indexação (1998) – *SIPORbase: sistema de indexação em português: manual*. 3ª ed. rev. e aumentada. Lisboa: BN, 1998. ISBN 972-565-154-5.

permite a sua gestão de um modo não possível nos antigos sistemas manuais e uniformizam a sua aplicação a nível quer da indexação, quer da pesquisa.

4.1.3 Outros sistemas

Para além dos sistemas de classificação universais, de estrutura hierárquica, e dos sistemas de indexação estruturados em tesouros e listas de cabeçalhos de assunto, desenvolveram-se outros instrumentos de apoio à análise e representação temáticas, com base em linguagens sintagmáticas que, possuindo um léxico tal como os tesouros, comportavam igualmente uma sintaxe que as tornava mais expressivas e precisas, mas também mais complexas – e é talvez esta complexidade uma das razões que as tornou menos populares que a linguagem combinatória dos tesouros⁷⁰. Estas linguagens aproximam-se das classificações por facetas, que se desenvolveram a partir dos anos 30 do séc. XX e que têm, entre as mais conhecidas, a *Colon Classification* de Ranganathan (1933). São também classificações deste tipo a *Bibliographic Classification* (BC), de Henry Evelyn Bliss, também conhecida como *Bliss Classification* (cuja tabela completa foi editada pela primeira vez entre 1940 e 1953) e o *Broad System of Ordering – BSO*⁷¹, desenvolvido pela FID, sob os auspícios da UNISIST, a partir de 1972 e cuja primeira versão provisória saiu em 1978. Embora inicialmente a intenção desta última iniciativa fosse criar uma linguagem intermediária para funcionar entre línguas (linguagens naturais) distintas (por exemplo, língua alemã - “linguagem de comutação” – língua francesa), que se aplicasse a todas as publicações para indicar as áreas temáticas por elas cobertas, o resultado acabou por ser um novo esquema de classificação geral (Rowley, 1987, p. 201).

As classificações por facetas são também chamadas analítico-sintéticas, uma vez que a composição da cadeia de representação resulta de dois processos: a análise (que decompõe o assunto nos seus conceitos básicos) e a síntese (que

⁷⁰ São exemplo de linguagens sintagmáticas o SYNTOL (*SYNTagmatic Organization Language*, 1960) e o PRECIS (*Preserved Context Indexing System*, 2ª ed.1984).

⁷¹ Também conhecido por *Standard Reference Code*.

combina os conceitos relevantes para descrever o assunto do item em causa). Uma faceta é um conjunto de “aspectos, propriedades e características de uma classe ou assunto específico, que é claramente definido, mutuamente exclusivo e colectivamente exaustivo” (Maple, 1997) e o tipo de classificação que lhe está ligado permite a combinação livre de conceitos de modo a representar um assunto mais complexo, pelo que se pode considerar um sistema dinâmico e flexível. A classificação *Colon* (assim designada por utilizar como símbolo os dois pontos, além de outros sinais de pontuação) começa por ter 42 classes principais, representativas das áreas do conhecimento. Cada classe é analisada e decomposta nos seus elementos básicos, agrupados por atributos comuns que se designam “facetadas”. Ranganathan reuniu essas facetadas em cinco grupos, a que chamou “categorias fundamentais”: personalidade, matéria, energia, espaço e tempo. Há facetadas que são comuns a todas as classes (as “*common isolates*”), como a forma ou a língua. As notações (compostas por letras e números) são usadas para representar as facetadas, a pontuação é usada para indicar a natureza e tipo das facetadas que se lhe seguem. A título de exemplo⁷², a classificação

L, 45; 421 : 6; 253 : f.44'N5

significa:

Medicina, Pulmões; Tuberculose : Tratamento; Raio X : Investigação.
Índia'1950

Concebida na Índia, a Classificação *Colon* (CC) não foi muito utilizada nos países ocidentais⁷³ mas é importante pelo que representa de tentativa de melhorar as formas de representação de assuntos, procurando resolver alguns dos problemas que os sistemas anteriores, nomeadamente as classificações universais, revelavam, para além de ter inspirado outros sistemas, como a classificação do catálogo britânico de música (a *British Catalogue of Music*

⁷² In Ranganathan: *ahead of his century*. [Em linha]. [Consult. 27 Out. 2004]. Disponível na WWW: <URL: <http://www.slais.ubc.ca/courses/libr517/winter2000/Group7/facet.htm>>.

⁷³ Contudo, B.C. Vickery publicou, em 1963, um guia para a construção de uma classificação por facetadas (*La classification à facettes : guide pour la construction et l'utilisation de schémas spéciaux*. Paris: Gauthier-Villars, 1963) e este tipo de classificação foi seguido pela CIS – *Classificação Internacional sobre a Segurança*, adoptado pelo Gabinete Internacional do Trabalho e pelo Sistema Internacional de Classificação *SfB*, para classificação de documentos relativos à construção, adoptado pelo respectivo ramo profissional e por vários países, entre os quais França.

Classification), desenvolvido por Eric Coates para a Bibliografia Nacional Britânica⁷⁴, e de ter contribuído para a actualização de sistemas de classificação já existentes (a CDU retomou as facetas “tempo” e “espaço”, representativas de noções de carácter universal, sob a forma de subdivisões comuns). Na CC, a composição da cadeia que representa o assunto utiliza uma sintaxe rigorosa e segue uma lógica que é a seguinte: a classe principal do assunto, a sequência PMEST (P- personalidade, M - matéria, E - energia, S - espaço e T – tempo), que equivale às facetas consideradas por Ranganathan “categorias fundamentais” e, eventualmente, outras indicações permitidas pelo sistema. Um aspecto positivo desta classificação é o facto de eliminar à partida a polissemia (um problema que as linguagens documentais, como linguagens controladas que são, devem resolver), já que as facetas têm também como objectivo reflectir o ponto de vista sob o qual os termos que integram a cadeia são considerados⁷⁵. Do ponto de vista estrutural, trata-se de uma linguagem documental que não está estabelecida à partida, mas mantém as divisões “tradicionais” do conhecimento herdadas da lógica aristotélica, pelo que a cadeia que representa o assunto está sempre ligada a uma classe principal. Contudo, o facto de não estar obrigada a uma divisão hierárquica existente *a priori*, torna-a, segundo Holzem (1999, p. 92), “mais apta a considerar a evolução disciplinar, assim como as relações interdisciplinares”. Por outro lado, a CC apresenta igualmente algumas analogias com os tesouros, se compararmos as facetas com o processo de decomposição da linguagem natural do utilizador em linguagem documental. Na elaboração de tesouros, as facetas são por vezes usadas nas relações hierárquicas entre os descritores, pois possibilitam a poli-hierarquia e permitem remeter para termos relacionados⁷⁶. E, embora a filosofia, a natureza, a estrutura e o desenvolvimento das listas de cabeçalhos de assunto diverjam da classificação por facetas, podemos considerar que os princípios subjacentes à CC foram também fonte de inspiração de alguns dos princípios orientadores para o

⁷⁴ Esta classificação foi usada para o catálogo de música entre 1957 e 1982, data a partir da qual influenciou as revisões da classe 780 da Classificação Decimal de Dewey e a Classificação de Bliss (Maple, 1997).

⁷⁵ Este aspecto (o ponto de vista sob o qual um assunto é abordado) é importante na actual concepção de ontologias.

⁷⁶ Estes tesouros designam-se por vezes “tesouros de facetas” e são exemplo de sistemas deste tipo o já referido sistema PRECIS, os MeSH (*Medical Subject Headings*) e o AAT (*Art and Architecture Thesaurus*).

estabelecimento de cabeçalhos de assunto, como os LCSH. Recentemente (2002-2003), o projecto FAKTS (*Facet Analytical Theory in Managing Knowledge Structure*), financiado pela AHRB (*Arts and Humanities Research Board*) – *Innovation Awards Scheme, 2001* – estudou a viabilidade de utilizar a teoria analítica das facetas para desenvolver uma estrutura de conhecimento que servisse para o ambiente digital e desenvolveu um protótipo⁷⁷ para a área das humanidades (uma ferramenta designada FAT-HUM), para o qual utilizou a análise de facetas da BC 2, da BSO e da CDU.

4.2 O acesso por assunto nos catálogos colectivos ou bases de dados cooperativas

Um dos problemas que se coloca ao acesso à informação existente em bases de dados de tipo cooperativo ou em rede, seja ela de carácter multidisciplinar ou especializado, tem a ver com a recuperação por assuntos. Mais uma vez se colocam, com grande acuidade, as questões da heterogeneidade. Aos inúmeros sistemas de representação de assuntos utilizados pelas bibliotecas e outros serviços de informação e à diversidade de critérios subjacentes à sua aplicação, junta-se ainda a incontornável questão do nível de profundidade da análise e representação dos assuntos, dependente das políticas de indexação adoptadas pelos serviços, assim como a irregularidade de tratamento, isto é, o facto de existirem documentos que são objecto de análise e representação de conteúdo e outros que não o são. Estas são algumas das questões que dificultam a consistência dos catálogos colectivos e impedem uma completa, correcta e eficaz recuperação da informação que contêm. A adopção de uma linguagem documental comum a todos os serviços de informação que participam nestas bases de dados pode aparecer como uma solução possível, embora apresente problemas reais difíceis de ultrapassar. Se olharmos para o exemplo da PORBASE – Base Nacional de Dados Bibliográficos, verificamos que a CDU (Classificação Decimal Universal) foi aí adoptada como linguagem comum pela comunidade das bibliotecas cooperantes, embora muitas delas nunca a tivessem utilizado na descrição de conteúdo da sua documentação e o passassem a fazer apenas para fins de cooperação – a obrigação de preencher o campo da CDU nos registos bibliográficos enviados para

⁷⁷ <URL: <http://www.ucl.ac.uk/fatks/prototype.htm>>.

integração na base de dados provoca resistências em algumas bibliotecas que não reconhecem a vantagem de utilizar uma tal linguagem além da que habitualmente utilizam e afasta mesmo algumas da cooperação⁷⁸. Outra solução tem sido a tentativa de compatibilizar as linguagens documentais usadas pela comunidade das bibliotecas, o que tem conduzido a iniciativas que procuram estabelecer as equivalências possíveis entre linguagens documentais de vários tipos. Estes mapeamentos podem contemplar apenas tesouros, tesouros e outros instrumentos de indexação terminológica, apenas classificações, ou classificações e linguagens terminológicas. Em 2002, Lois Mai Chan e Marcia Lei Zeng apresentaram, na 68ª Conferência Geral da IFLA, uma comunicação sobre a interoperabilidade entre vocabulários de assunto e esquemas de organização do conhecimento⁷⁹ em que apresentavam uma análise dos métodos mais utilizados para atingir e melhorar essa interoperabilidade. A partir do estudo de várias experiências e projectos (interoperabilidade entre vocabulários controlados na mesma língua, entre múltiplos vocabulários de assunto em diferentes línguas e sistemas de classificação, entre um vocabulário controlado e um sistema de classificação universal, entre sistemas de classificação ou recorrendo a novos sistemas para diferentes línguas), dividiram em dois grupos os resultados obtidos:

- 1 - o dos métodos usados para alcançar e melhorar a interoperabilidade;
- 2 - o dos métodos utilizados para armazenar e gerir as ligações.

Em relação ao primeiro grupo, concluíram que os métodos mais seguidos haviam sido os seguintes:

- Derivação/Modelação, em que, como modelo ou ponto de partida, um vocabulário mais simples ou mais especializado se desenvolve juntamente com um vocabulário mais abrangente;
- Tradução/Adaptação, em que os termos do vocabulário são traduzidos, com ou sem adaptações, de um vocabulário já existente noutra língua;

⁷⁸ Na nossa opinião, reflecte-se aqui a constatação de Aida Slavic de que muitas bibliotecas, serviços bibliográficos e sistemas não exploram completamente a classificação e de que existem ainda muitas confusões sobre o que se pode conseguir com sistemas como a CDU (2003, p. 2).

⁷⁹ Ensuring interoperability among subject vocabularies and knowledge organization schemes : a methodological analysis. IFLA COUNCIL AND GENERAL CONFERENCE, 68, Glasgow, 2002 – *Proceedings*. [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla68/papers/008-122e.pdf> >

- Mapeamento (intelectual), em que se desenvolve um sistema de mapeamento que consiste basicamente em estabelecer equivalência entre termos de diferentes vocabulários, ou entre termos verbais e notações;
- Mapeamento (com apoio de computador), em que o sistema de mapeamento assenta parcial ou substancialmente em tecnologia informática;
- Ligação, em que os termos de uma lista se ligam a outros termos não equivalentes a nível conceptual mas relacionados a nível linguístico, de modo a otimizar os resultados da recuperação;
- Comutação, em que uma linguagem ou esquema comutador serve de intermediário entre termos equivalentes em diferentes vocabulários.

Já em relação ao segundo grupo (métodos utilizados para armazenar e gerir as ligações), identificaram três métodos:

- Registos de autoridade, utilizando campos específicos nos formatos de autoridade para armazenar ligações;
- Concordâncias;
- Rede semântica, que consiste numa estrutura organizada, que serve de “coluna vertebral”; cada unidade da rede representa um conceito à volta do qual se identifica e armazena um agregado (“*cluster*”) de termos equivalentes de diferentes vocabulários.⁸⁰

Martin Doerr (2001), por seu turno, estudou as diferentes alternativas à compatibilização de recursos terminológicos, começando por analisar a tentativa de unificar tesouros (ou outras linguagens terminológicas), fundindo-as – apontava como exemplo o *Unified Medical Language System (UMLS)*, que fundiu conceitos de 50 fontes diferentes num metatesouro, mantendo ligações às suas fontes originais. Esta solução não se revelou muito adequada, já que diferenças semânticas dos termos e das relações hierárquicas, assim como a sobreposição de termos, podem tornar a fusão impossível; além disso, esta solução revelava-se igualmente muito difícil de implementar, dada a quantidade de sistemas que utilizam as linguagens e as dificuldades de fazer migrar para

⁸⁰ Chan e Zeng (2002, p. 3-4).

esses sistemas a terminologia e todo o sistema de relações. Doerr analisou também a experiência de harmonização de vocabulários através da correlação ou da federação de tesouros ou de listas de cabeçalhos de assunto, desenvolvidas por meio de mapeamentos. Referiu alguns aspectos da heterogeneidade detectada no mapeamento de tesouros num mesmo domínio, traduzida nas diferenças existentes a nível do uso das palavras (devido à existência de diferentes linguagens naturais, ao nível de linguagem escolhido, a decisões sobre a selecção de descritores não justificadas semanticamente e ao grau de pós ou pré-coordenação da terminologia), da cobertura (devido a diferentes estádios de desenvolvimento dos tesouros, a diferentes finalidades e a diferentes necessidades dos utilizadores), da semântica (devido a diferentes conceptualizações) e das relações semânticas (frequentemente devido à aplicação de mono-hierarquias, mas também devido a diferentes aspectos de categorização). Estas diferenças, detectadas entre tesouros que procuram cobrir apenas uma determinada área temática, podem atingir níveis que tornem demasiado morosas ou mesmo inoportáveis as tarefas de mapeamento, se alargarmos o âmbito desta heterogeneidade aos vários domínios temáticos e aos diferentes sistemas que os procuram representar. Neste ponto, pode ser-nos útil olhar para o modo como os assuntos e a sua recuperação têm sido encarados no mais vasto mundo da informação que é a *Web*.

4.3 A recuperação por assuntos na *Web*

Num quadro genealógico das linguagens documentais (que se pode ver no anexo 2), Menon (2004) procura demonstrar que, na designada “nova geração de sistemas de organização do conhecimento”, podemos reconhecer quase todas as componentes das linguagens documentais tradicionais, embora essa herança nem sempre seja reivindicada. O facto de as linguagens documentais, registadas sobretudo em tabelas ou em esquemas de classificação e em tesouros, terem sobrevivido até hoje, justifica-se pela sua adaptabilidade às novas situações que, ao longo do tempo, foram surgindo. Daí que, como dizia Jacques Chaumier no final de 2000, “*les langages documentaires n’ont pas bougé*”⁸¹.

⁸¹ *Archimag*, n.º 139 (Nov. 2000), citado por Menon (2004).

As novas necessidades criadas com o advento da *Web* vieram obrigar, por um lado, a uma reflexão sobre as características da informação assim disponibilizada e, por outro, a um renovado esforço de selecção a nível dos conteúdos a integrar nas bases de dados e a dar corpo às bibliotecas digitais. O aumento da disponibilidade de textos integrais acessíveis a partir de qualquer ponto do globo veio abrir caminho ao desenvolvimento de formas de recuperação temática assentes na pesquisa em texto livre, embora tenham igualmente surgido experiências mais elaboradas e complexas, mas também mais rigorosas, de aceder à informação neles contida. Estas experiências são igualmente subsidiárias dos desenvolvimentos da terminologia enquanto disciplina científica, do desenvolvimento de bases de dados terminológicos e da investigação na área da linguística computacional, da pesquisa e recuperação multilíngues (como, apenas a título de exemplo, a que tem vindo a ser desenvolvida no âmbito do CLEF – *Cross-Language Evaluating Forum*⁸²). Paralelamente à actualização e ao aperfeiçoamento dos tesouros e dos sistemas de classificação usados nas bibliotecas, têm vindo a testar-se novos instrumentos para permitir a organização e representação dos conteúdos temáticos dos objectos digitais e do ambiente *web*, concebidos tendo como ponto de partida a complementaridade entre linguagens controladas e linguagem natural. Alguns são novos “tesouros conceptuais”, que se usam durante o processo de indexação “como ajuda na identificação dos conceitos, e no processo de recuperação, como fonte de novos termos que identifiquem conceitos e aumentem a precisão das pesquisas” (Lopez Alonso, 2001, p. 12). Procura-se, com estes instrumentos, uma integração da linguagem natural, usada pelo utilizador, com as linguagens documentais que, cada vez mais, se conjugam com a terminologia, já que os textos técnicos estão armazenados em suporte digital e podem converter-se num formato adequado para a análise terminológica de *corpora* especializados (Lopez Alonso, 2001, p. 13).

A investigação na área da recuperação baseada em linguagem natural, embora com novos contornos que advêm do desenvolvimento da Internet e da *Web* e das tecnologias que lhe estão associadas, integra-se no conjunto das

⁸² <URL: <http://clef.isti.cnr.it/>>.

experiências encetadas e desenvolvidas na 2ª metade do séc. XX que procuravam desenvolver a indexação automática de documentos em texto integral e se baseavam, em parte, na utilização de métodos estatísticos ligados às ocorrências de uma mesma palavra num texto, que a podiam tornar suficientemente relevante para ser indexada e servir depois como palavra-chave na recuperação – a maior ou menor frequência de um termo num texto ditava assim a sua selecção ou a sua exclusão como “termo de indexação”. Embora podendo resolver o problema do tratamento de grandes volumes de informação de uma forma rápida e praticamente sem intervenção humana, estes processos têm-se revelado insuficientes enquanto formas de analisar assuntos e representar o conhecimento, pois o facto de uma palavra só aparecer uma vez num texto não implica que ela não corresponda a um conceito a reter, tal como a situação inversa é igualmente válida. Além disso, questões como a sinonímia, a homografia e a polissemia não são devidamente consideradas, pelo que tais sistemas teriam de ser mais aperfeiçoados e desenvolvidos, no sentido da referida complementaridade entre linguagem natural e linguagem documental.

Mas uma questão fulcral da recuperação na Internet prende-se com os instrumentos desenvolvidos para facilitar o acesso a todo o acervo informativo disponível, dos motores de busca às formas mais ou menos estruturadas de organização da informação.

4.3.1 Os portais temáticos (“Subject gateways”)

Os portais temáticos surgiram como resultado de uma reflexão aberta e partilhada sobre a selecção da informação a circular na rede e reflectem o trabalho conjunto de diferentes entidades, concretizado em diversos projectos desenvolvidos com o objectivo de criar formas unívocas de aceder, de um modo consistente, à informação existente na rede sobre determinados assuntos. Alguns desses projectos têm continuado a desenvolver-se após a

fase piloto, outros ficaram-se praticamente pela fase experimental⁸³, pois a manutenção de portais temáticos requer uma afectação de recursos bastante significativa – mediante o volume de informação nova que diariamente surge na rede, nem sempre existem condições e meios para que a sua actualização seja constante. Na verdade, embora possamos encontrar uma certa semelhança entre o trabalho tradicionalmente desenvolvido pelas bibliotecas na construção e desenvolvimento dos seus catálogos e a criação e manutenção de portais temáticos, as diferenças são assinaláveis. Chapman, Day e Hiom (1998) compararam as práticas inerentes ao tradicional desenvolvimento dos catálogos de bibliotecas com a abordagem que é feita nos portais temáticos baseados na Internet, tomando como exemplo o SOSIG⁸⁴. Procurámos sintetizar essa comparação no seguinte quadro:

	Bibliotecas (catálogos tradicionais)	Portais temáticos
Seleção	<ul style="list-style-type: none"> • Separada da catalogação 	<ul style="list-style-type: none"> • Levada a cabo por especialistas de assunto e profissionais da informação, frequentemente também responsáveis pela catalogação dos recursos
Descrição	<ul style="list-style-type: none"> • Descrição dos itens de acordo com regras de catalogação 	<ul style="list-style-type: none"> • A descrição física (n.º de páginas, menção de ilustração, etc.) dos recursos da Internet não é necessária. • Há geralmente poucos dados bibliográficos disponíveis através do próprio recurso, o catalogador tem muitas vezes de investigar quem é o responsável pela informação. • A descrição pode permitir mais texto livre
Cabeçalhos	<ul style="list-style-type: none"> • Pontos de acesso para uma entrada do catálogo (autor, assunto, classificação...) 	<ul style="list-style-type: none"> • Pontos de acesso: título, autor, categoria (tipo de recurso), assunto, palavras de descrição...

⁸³ Está neste caso o portal temático ADAM (*Art, Design, Architecture & Media Information Gateway*), suspenso em 2003. Um novo serviço, o *Arts and Creative Industries (ACI) Hub*, daria continuidade ao seu papel de portal temático.

⁸⁴ *Social Science Information Gateway* (<URL: <http://www.sosig.ac.uk/>>).

Controlo de autoridade	<ul style="list-style-type: none"> • Estabelece as formas preferenciais de nomes e termos 	<ul style="list-style-type: none"> • Contribui para a consistência num serviço individual e facilita o processo de pesquisa cruzada com outros portais e catálogos
Existências e informação de acesso	<ul style="list-style-type: none"> • As questões sobre o acesso estão frequentemente relacionadas com a informação sobre existências 	<ul style="list-style-type: none"> • Os portais estão mais ligados ao acesso do que às existências. • A natureza volátil da Internet requer que os registos estejam constantemente a ser controlados, com o objectivo de verificar se as descrições dos recursos são ainda correntes e se as ligações ainda funcionam.

O fundamento dos portais temáticos – fornecer um acesso fácil a informação de qualidade sobre determinado(s) assunto(s) existente(s) na Internet - coloca um enfoque muito especial na possibilidade de se aceder aos conteúdos temáticos dos recursos que descrevem, pelo que a sua indexação ou classificação é considerada um aspecto essencial do serviço que disponibilizam⁸⁵. Estes portais optam, em geral, por utilizar sistemas de organização e representação do conhecimento já existentes, sejam as grandes classificações enciclopédicas como a CDU (adoptada pelo já referido SOSIG), a CDD (usada, por exemplo, pelo PSIGATE⁸⁶) e a LCC (estrutura do VETGATE⁸⁷) ou sistemas de classificação especializados, como a ACM *Computing Classification*⁸⁸ ou a *Mathematics Subject Classification*⁸⁹ (seguidas pelo EEVL⁹⁰), para além de usarem também, por vezes de forma

⁸⁵ Um portal de referência como o EELS (*Engineering E-Library, Sweden*), ao suspender a indexação "manual" e ao "congelar" a versão do serviço disponível na Internet, avisou no seu site que as bibliotecas das universidades de tecnologia suecas não garantiriam, a partir de então, a qualidade do seu conteúdo, à semelhança do que acontecia com o índice *All Engineering*. Os utilizadores eram ainda informados de que as bibliotecas do Real Instituto de Tecnologia e da Universidade de Lund, na Suécia, estavam a trabalhar no sentido da criação de um novo serviço, baseado no *All Engineering* original e que consistiria nos registos seleccionados automaticamente, pela sua relevância, de acordo com o *Engineering Index Thesaurus* (Ei) Cf. *Notice - EELS has closed down!*. [Em linha]. [Consult. 5 Mar. 2005]. Disponível em: <URL:<http://eels.lib.lu.se/>>.

⁸⁶ <URL: <http://www.psigate.ac.uk/>>

⁸⁷ <URL: <http://vetgate.ac.uk/>>

⁸⁸ <URL: <http://www.acm.org/class/>>

⁸⁹ <URL: <http://www.ams.org/msc/>>

⁹⁰ <URL: <http://www.eevl.ac.uk/>>

complementar, tesouros ou outros instrumentos terminológicos (como, por exemplo, o tesouro Ei - *Engineering Index Thesaurus*, utilizado também pelo EEVL e pelo EELS)⁹¹. Para os participantes no projecto DESIRE, a utilização de classificações, em particular das grandes classificações gerais e universais, para organizar e recuperar recursos da Internet apresentava vantagens e desvantagens. Quanto às primeiras, consistiam na facilidade de “*browsing*”, na possibilidade de alargar ou restringir o âmbito da pesquisa, no contexto, no potencial para o acesso multilíngue, na facilidade em manipular e utilizar apenas parcialmente as suas bases de dados, no facto de serem esquemas resultantes de consenso, de não correrem o risco de obsolescência e de serem largamente conhecidas e utilizadas. Quanto às desvantagens, elas radicavam principalmente na divisão de colecções lógicas de materiais (o que se ultrapassaria com um bom sistema de referências cruzadas), na subdivisão ilógica de algumas classes e na lenta assimilação de novas áreas de interesse (Koch et al., 1997, p.2).

Mais adiante analisaremos a opção seguida pelo Projecto Renardus, que procurou facultar o acesso simultâneo, através de uma única chave de pesquisa, a todo um conjunto de informação diversa existente nos portais temáticos, indexada ou classificada com base em diferentes sistemas e diferentes linguagens.

4.3.2 As ontologias

Especialmente a partir da última década do séc. XX, paralelamente ao aumento de disponibilidade de recursos digitais e às iniciativas da *web* semântica, desenvolveram-se as ontologias, formas de representação de um determinado domínio do conhecimento, que permitem o estabelecimento de relações e apresentam suficiente flexibilidade para poderem ser partilhadas e reutilizadas. Podem conter termos descritivos variáveis e, acima de tudo, podem exprimir toda a estrutura semântica e todas as relações numa linguagem compreendida

⁹¹ Um levantamento dos vários esquemas de classificação, das suas características e adaptabilidade à descrição e descoberta de recursos da Internet foi feito no âmbito do Projecto DESIRE (*Development of a a European Service for Information on Research and Education*). Veja-se Koch et al. (1997).

pelo computador. Têm, no fundo, semelhanças com as estruturas convencionais usadas para organizar e representar o conhecimento (as classificações e os tesouros), embora apresentem igualmente diferenças que advêm sobretudo do facto de serem concebidas para uma utilização em ambiente informático e não num ambiente puramente humano. As semelhanças existentes levam a que seja reconhecida a vantagem de utilizar, como ponto de partida para o estabelecimento de ontologias, todo o conhecimento e esforço intelectual já aplicado na construção de linguagens documentais materializadas em classificações e, sobretudo, em tesouros⁹². Para alguns dos intervenientes na transformação do tesouro *Agrovoc* numa ontologia, as diferenças entre tesouros e ontologias revelam-se a nível dos utilizadores (humanos *versus* humanos / agentes de *software*), da semântica (ausência frequente de uma especificação formal dos conceitos *versus* semântica claramente explícita que permite uma interpretação não ambígua dos termos, essencial a uma utilização por computadores) e a nível de representação do conhecimento em suporte informático (limitada ou inexistente *versus* explícita e formalizada). Mas reconhecem que ambos (tesouros e ontologias) permitem a representação de formas de entendimento partilhado de um domínio, facilitando uma comunicação eficiente, são baseados em conceitos e respeitam a terminologia adequada para os representar, usam hierarquias para agrupar termos em categorias e subcategorias e são aplicados à catalogação e organização dos recursos de informação (Fisseha, Liang e Keizer, 2003).

Embora não exista ainda uma definição única universalmente aceite para a ontologia (Crofts, Le Boeuf e Artur, 2002), Natalya Noy e Deborah McGuinness propõem, para os objectivos do seu guia para a criação de uma primeira ontologia, a seguinte definição: “uma ontologia é uma descrição formal explícita de conceitos num domínio do discurso (classes - por vezes chamadas conceitos), de propriedades de cada um desses conceitos descrevendo várias das suas características e atributos (“*slots*” - por vezes designados funções ou

⁹² São exemplos os trabalhos desenvolvidos a partir do tesouro *Agrovoc*, ou do vocabulário controlado do GEM (*Gateway to Educational Materials*) (Fisseha, Liang e Keizer, 2003; Qin e Paling, 2001).

propriedades) e de restrições aos “*slots*” (facetas - por vezes chamadas restrições de função)” (Noy e McGuinness, [2001], p.3). As classes são consideradas o foco da maioria das ontologias e podem ter subclasses que representem conceitos mais específicos. Os passos a seguir na metodologia de desenvolvimento de uma ontologia, ainda segundo os mesmos autores, são os seguintes:

- Determinar o domínio e finalidade da ontologia
- Considerar a reutilização de outras ontologias
- Enumerar termos importantes na ontologia
- Definir as classes e a hierarquia de cada classe
- Definir as propriedades das classes – “*slots*”
- Definir as facet
- Criar instâncias

Os representantes do ELAG (*European Library Automation Group*) no seminário sobre a *web* semântica e as bibliotecas realizado em Roma em 2002, além de analisarem o significado das ontologias apresentaram um modelo de referência para a sua construção: o CRM (*Conceptual Reference Model*) do ICOM / CIDOC, uma ontologia desenvolvida para o sector do património cultural. Segundo eles, uma ontologia pode definir-se como “uma representação formal e explícita de uma conceptualização específica”: numa ontologia existe uma representação formal que visa precisar, explicitar e evitar ambiguidade, a ontologia descreve as entidades e conceitos relevantes para um domínio específico e corporiza um particular ponto de vista de um domínio que pode ser descrito de diferentes modos por diferentes ontologias (Crofts, Le Boeuf e Artur, 2002). Desenvolvem igualmente, a par do modelo CRM, algumas ideias sobre o modo de construir ontologias e os vários passos a seguir: definir o objectivo e finalidade da ontologia, identificar as entidades específicas do domínio, organizá-las numa hierarquia e defini-las, identificar as suas propriedades, identificar as relações que estabelecem entre si, descrevê-las e defini-las, refinar e expandir a ontologia (no anexo 3 podemos ver como os autores, tomando como exemplo um acto do quotidiano – preparar uma chávena de chá – apresentam um método para a construção de ontologias

baseado, embora de um modo simplificado, nos mesmos passos lógicos envolvidos na criação do CRM; os próprios afirmam ter escolhido um exemplo divertido, para não ser levado muito a sério).

Na realidade, a construção de uma ontologia envolve igualmente, como os tesouros, um elevado investimento intelectual, que se pode rentabilizar pelo facto de permitir um tratamento automatizado e tão rigoroso quanto possível do volume de informação acessível em linha e que contribui para uma análise mais detalhada do domínio sobre o qual se debruça, avivando “esse sentido de descoberta [que] é, afinal, uma das razões principais de fazer uma ontologia” (Crofts, Le Boeuf e Artur, 2002).

O WWW Consortium, (W3C) através do *Web Ontology Working Group*, inserido na *W3C Semantic Web Activity*, publicou, em Fevereiro de 2004, um conjunto de recomendações para as OWL (*Web Ontology Languages*), dividido em seis partes. Numa delas (*Use cases and requirements*) são apresentados os requisitos para uma linguagem de ontologia *web*, considerando as ontologias como recursos distintos, com os seus próprios identificadores únicos. São eles:

- Referenciação não ambígua de conceitos através de URIs;
- Extensão explícita da ontologia, para reutilizar conceitos aquando da adição de novas classes e propriedades (a extensão deverá ser uma relação transitiva: se a ontologia A expande a ontologia B e esta a C, então a ontologia A expande também a ontologia C);
- Envio para ontologias – os recursos devem ter a capacidade de remeter para ontologias específicas;
- Metadados da ontologia (adoptados ou não do conjunto de elementos Dublin Core);
- Informação que distinga as várias versões;
- Definição das classes (a linguagem tem de ter a capacidade de exprimir definições complexas de classes, incluindo a divisão em subclasses e combinações booleanas de classes);
- Definição das propriedades (incluindo subpropriedades, restrições de domínio, transitividade, propriedades inversas);

- Tipos de dados – a linguagem deve fornecer um conjunto de tipos de dados *standard* (podem basear-se no esquema XML de tipos de dados);
- Equivalência de classes e propriedades – possibilidade de indicar a equivalência de duas classes ou de duas propriedades;
- Equivalência individual – possibilidade de indicar que pares de indicadores representam a mesma instância (é provável que a mesma instância possua diferentes identificadores - o uso de um URL *standard* não resolve a situação, já que algumas instâncias podem ter múltiplos URLs⁹³);
- Ligar informação a explicações – a linguagem deve possuir um mecanismo que permita “etiquetar” as explicações com informação adicional (fonte, data, nível de confiança, etc.);
- Classes como instâncias – a linguagem deve ter a possibilidade de tratar classes como instâncias, já que o mesmo conceito pode ser uma classe ou uma instância, dependendo da perspectiva do utilizador (por exemplo, a classe “Orangotango” pode ter como instâncias animais individuais, embora ela própria seja uma instância da classe “Espécies”);
- Restrições numéricas – a linguagem pode especificar restrições às propriedades (indicando o número máximo e mínimo de indivíduos com que qualquer instância pode estar relacionada, por via da propriedade especificada);
- Sintaxe XML – necessária para garantir a compatibilidade;
- Apresentação ao utilizador – a linguagem deverá permitir a especificação de múltiplas apresentações alternativas para os recursos especificados por uma ontologia (para se ver a ontologia em diferentes línguas (línguas naturais), por exemplo);
- Modelo de caracteres – a linguagem deverá permitir o uso de conjuntos de caracteres multilíngues;
- Unicidade de cadeias de sequência UNICODE.

O cumprimento destes requisitos básicos deverá garantir a evolução das ontologias, a sua partilha e interoperabilidade, o equilíbrio da expressividade e

⁹³ O caso, por exemplo, de pessoas que têm páginas *web* de trabalho e particulares.

escalabilidade⁹⁴, a detecção de inconsistências, a compatibilidade com outros *standards*, a internacionalização e a facilidade de utilização, independentemente da necessidade de se definirem outras características consideradas úteis na construção da linguagem.

Estas recomendações aproximam-se, do ponto de vista da intenção, das normas desenvolvidas no passado (e ainda válidas, embora se defenda a necessidade da sua revisão à luz dos desenvolvimentos que temos vindo a referir) para a construção de tesauros monolíngues e multilíngues. O CIDOC *Conceptual Reference Model*, atrás referido, está mesmo a ser transformado em norma pela ISO⁹⁵, podendo tornar-se num importante instrumento de trabalho para a cooperação entre bibliotecas, arquivos e museus, já que se trata de “uma ontologia formal que tem como intuito facilitar a integração, mediação e intercâmbio de informação heterogénea na área do património cultural” (ICOM, 2004, p. i). Culminando a actividade de normalização desenvolvida durante mais de uma década pelo CIDOC (*International Committee for Documentation*) do ICOM (*International Council of Museums*)⁹⁶, o seu objectivo é definir uma estrutura que permita a interoperabilidade semântica e não tratar os aspectos ligados à terminologia em si⁹⁷. Tem sido expressa como um modelo semântico orientado para objectos e pode, por isso, ser implementada em qualquer esquema relacional ou orientado para objectos e as suas instâncias podem ser codificadas em RDF, XML, DAML-OIL⁹⁸ e OWL, entre outras.

⁹⁴ Escalabilidade: “capacidade de expansão. Implica alterações mínimas em procedimentos correntes de forma a permitir o crescimento” (Sousa, 2001, p. 227).

⁹⁵ ISO/DIS 21127. 2005, *Information and documentation : A reference ontology for the interchange of cultural heritage information*. O trabalho de adaptação a norma foi desenvolvido no âmbito do grupo de trabalho 9 da Subcomissão 4 – Interoperabilidade técnica – da Comissão Técnica 46 – Informação e Documentação – da ISO.

⁹⁶ De que podemos destacar as *International guidelines for museum object information : the CIDOC information categories* [Em linha]. [Consult. 9 Fev. 2005]. Disponível na WWW: <URL: <http://www.willpowerinfo.myby.co.uk/cidoc/guide/guideint.htm>>.

⁹⁷ Para informação sobre o CIDOC CRM, veja-se <URL: <http://cidoc.ics.forth.gr/>>

⁹⁸ DAML – *DARPA Agent Markup Language* (<URL: <http://www.daml.org/>>); OIL - *Ontology Inference Layer* (<URL: <http://www.ontoknowledge.org/oil/>>)

4.3.3 Outras iniciativas para a organização do conhecimento na *Web*

Paralelamente ao desenvolvimento específico de ontologias, outras iniciativas têm procurado, desde os últimos anos da década de 90, adequar os tesouros e outros tipos de sistemas de organização do conhecimento à realidade da *web* e das bibliotecas digitais, desenvolvendo formatos normalizados e protocolos que facilitem a interoperabilidade a nível de acesso, recuperação e troca de informação. Surgiram assim os primeiros passos de um formato de troca RDF (*RDF Interchange Format*) e de ferramentas Java para tesouros em linha, que “empacota” os dados do tesouro em RDF e emprega um protocolo para tesouros HTTP (Miles et al., 2003), com o objectivo de os desenvolver como serviços *web* no projecto SWAD-EUROPE⁹⁹. Também um outro projecto, o OCLC *Terminology Services Project*, emprega o protocolo Z39.50 (SRUW) com um perfil modificado Zthes, para disponibilizar e expandir os recursos terminológicos através de vários serviços *web*, incluindo o mapeamento automático entre diferentes recursos (DDC e LCSH) (Vizine-Goetz, 2003). O protocolo CERES, apresentado nos primeiros meses de 1999, destinava-se a facilitar a comunicação entre tesouros distribuídos, no âmbito de um projecto¹⁰⁰ que visava a construção integrada de um vocabulário controlado sobre ambiente, juntamente com as ferramentas que permitissem o seu uso para a criação de metadados e formulação de pesquisas (“*query construction*”), tanto isoladamente como em sistemas *web*. O CERES desenvolvia um protocolo HTTP usando um formato de representação de tesouros em RDF(XML) que seguia a norma NISO Z39.19¹⁰¹ (Binding e Tudhope, 2004). Sensivelmente na mesma altura, era apresentado o já atrás referido perfil Zthes do protocolo Z39.50, seguindo a norma ISO 2788. Embora parte da especificação diga respeito à representação dos registos das bases de dados de tesouros para implementação em Z39.50, o objectivo era criar um modelo suficientemente geral para poder ser usado por outros protocolos de comunicação, apresentando-se assim uma DTD (*Document Type Definition*) para tesouros

⁹⁹ <URL: <http://www.w3.org/2001/sw/Europe/>>

¹⁰⁰ O CERES/NBII *Thesaurus Partnership Project*, desenvolvido pelo *Californian Environmental Resources Evaluation System* (CERES) e pela *US Geological Survey Biological Resources Division* (USGB/BRD), com o objectivo de facilitar o acesso à informação sobre ambiente <URL: <http://ceres.ca.gov/thesaurus/>> e <URL: <http://ceres.ca.gov/index.html>>.

¹⁰¹ A norma NISO Z39.19 equivale à norma ISO 2788.

em XML. O perfil Zthes tem sido usado como parte do ZING (Z39.50 - *International: New Generation*), no protocolo SRW *Search/RetrieveWeb Service*). O conjunto de qualificadores Zthes para CQL (*Common Query Language*, usada pelo SRW) permite os seguintes serviços Zthes: pesquisar representações em tesouros para identificação de conceitos ou encadeamento de termos; encontrar ocorrências de palavras em qualquer ponto do registo do termo; pesquisar todos os termos mais específicos, o termo mais genérico, o termo preferencial, os termos não preferenciais e os termos relacionados de um dado conceito; pesquisar os “equivalentes linguísticos” de um conceito (Binding e Tudhope, 2004). Um outro protocolo para tesouros é o ADL (*Alexandria Digital Library*)¹⁰², baseado em HTTP e XML e que segue igualmente a norma NISO Z39.19. Este protocolo visa descarregar, interrogar (pesquisar) e navegar através de tesouros e foi desenvolvido em paralelo com um protocolo para “dicionário” de termos geográficos (“*gazetteer*”).

Binding e Tudhope, após analisarem os protocolos que acabámos de referir, concluem que a tendência actual das arquitecturas orientadas para serviços (SOA – *Service Oriented Architectures*) “traz oportunidades de se caminhar para uma separação clara entre as componentes da interface e as fontes de dados subjacentes, através de serviços *Web* adequados”, com as vantagens de “disseminação neutra, a nível de plataformas, dos conteúdos do tesouro, potenciando o esforço intelectual envolvido na compilação de tesouros - explorando representações comuns, etc.”. Defendem ainda que um protocolo para tesouros deverá incluir um serviço de expansão semântica e consideram que, estando formatos e protocolos intrinsecamente ligados, ambos devem evoluir no sentido de se atingirem *standards* comuns (Binding e Tudhope, 2004).

¹⁰² <URL: <http://www.alexandria.ucsb.edu/>>

5. Três abordagens complementares à questão do acesso por assunto

5.1 O Projecto RENARDUS – um exemplo de organização do conhecimento e de interoperabilidade entre diferentes sistemas de classificação e um modelo comum

O Projecto Renardus, financiado pela Comissão Europeia ao abrigo do Programa das Tecnologias da Sociedade de Informação do 5º Programa-quadro da União Europeia, desenvolveu-se entre Janeiro de 2000 e Junho de 2002 e reuniu 12 parceiros (bibliotecas nacionais, centros de investigação e portais de assunto) de 7 países (Dinamarca, Finlândia, Alemanha, Holanda, Suécia, França e Reino Unido), coordenados pela Biblioteca Nacional da Holanda¹⁰³. Surgiu na sequência de outros projectos europeus, como o atrás referido Projecto DESIRE, que desenvolvera soluções para a descrição e classificação automática de conteúdos, ao nível dos portais de assunto individuais, com base em esquemas de classificação já consolidados. Como objectivo, o Projecto RENARDUS visava permitir à comunidade académica e aos investigadores europeus (grupos-alvo do projecto) um acesso integrado, através de uma única interface, a serviços distribuídos baseados na Internet, no caso portais de assunto com qualidade controlada¹⁰⁴, usando para tal um mesmo perfil de metadados e um sistema de classificação comum ao qual todos os sistemas de classificação locais fossem mapeados. O perfil de

¹⁰³ Participaram no Projecto, entre Janeiro de 2000 e Junho de 2002: a *Koninklijke Bibliotheek* (KB), com responsabilidade de coordenação; a *Bibliothèque Nationale de France* (BNF), o *Center for Scientific Computing*, Finlândia (CSC), *Die Deutsche Bibliothek* (DDB), o *Finnish Virtual Library Project*, *Jyväskylä University Library*, Finlândia (JyU), o *Institute for Learning and Research Technology*, Universidade de Bristol, Reino Unido (ILRT), o *NetLab*, Universidade de Lund, Suécia (NetLab), a *Niedersächsische Staats- und Universitätsbibliothek*, Göttingen, Alemanha (SUB), o *Technical Knowledge Centre and Library of Denmark*, Dinamarca (DTV), o *UK Office for Library and Information Networking*, Universidade de Bath, Reino Unido (UKOLN), a *Viiikki Science Library*, Universidade de Helsinquia, Finlândia (ALUH), e o *Zentralstelle für Agrardokumentation Und-information*, Alemanha (ZADI).

¹⁰⁴ No âmbito do projecto, consideraram-se portais de assunto com qualidade controlada “serviços Internet que utilizam um conjunto medidas de qualidade para apoiar a descoberta sistemática de recursos e que desenvolvem um considerável esforço manual não só para assegurar uma selecção de recursos que vão ao encontro de critérios de qualidade, mas também para apresentar uma descrição rica desses recursos, com base em metadados normalizados. O controlo e a actualização regulares asseguram uma boa gestão das colecções. Um dos objectivos principais destes portais é permitir um acesso por assunto de alta qualidade através de recursos de indexação, usando vocabulários controlados e oferecendo uma estrutura de classificação profunda para navegação e pesquisa avançada.” (Koch, 2000)

metadados adoptado baseou-se no Dublin Core, com excepção do elemento “País” (veja-se o anexo 4), tendo os metadados locais, assim como os seus códigos de conteúdo, sido mapeados nesse perfil comum. Constituiu-se, assim, a infra-estrutura necessária à interoperabilidade entre as bases de dados participantes, pré-requisito essencial à efectivação da pesquisa cruzada. A escolha do conjunto de metadados Dublin Core foi justificada pela sua adequação à descoberta e à descrição de recursos *web*, sendo considerada uma estrutura simples, com capacidade para permitir a interoperabilidade semântica, reunindo consenso internacional, fornecendo uma alternativa aos modelos de descrição mais elaborados (como os formatos MARC completos) e possuindo suficiente flexibilidade para possibilitar a codificação de semânticas mais elaboradas¹⁰⁵. Além destas razões, o formato Dublin Core era suficientemente genérico para abarcar todas as áreas disciplinares cobertas pelos participantes no projecto e suficientemente geral para servir de formato normalizado ao qual os serviços pudessem facilmente mapear os seus próprios formatos (Peereboom, 2000, p.7-8). Contudo, para garantir a interoperabilidade foi necessário chegar a consenso entre os serviços que usavam já algum tipo de implementação Dublin Core e aqueles que usavam outros formatos. Mas o questionário enviado no início do projecto aos potenciais serviços participantes demonstrou que a maioria desejava que Renardus admitisse a semântica Dublin Core, assim como a sintaxe XML/RDF, embora estivesse disposta apenas a pequenos ajustamentos nos seus serviços para participar no projecto¹⁰⁶. Uma vez que alguns serviços não comportavam todos os

¹⁰⁵ A conciliação entre a exigência de simplicidade, por um lado, e a necessidade de maior complexidade que alguns implementadores do Dublin Core sentiam, foi ultrapassada pela criação de novos elementos, assim como de qualificadores locais (Peereboom, 2000, p. 7). Cf. anexo 4.

¹⁰⁶ Os portais de assunto que responderam aos primeiros questionários lançados pelo Projecto foram os seguintes: DutchESS (Holanda), NOVAGate (países nórdicos), EELS (Suécia), DEF *fagportal* (Dinamarca), DAINet (Alemanha), FVL (Finlândia), Les Signets (França), RDN (Reino Unido), DDB (Alemanha) and SSG-FI (Alemanha). As respostas que deram a um questionário elaborado pelo SUB conduziram ao desenvolvimento de vários modelos de dados, entre os quais o modelo de dados do sistema protótipo piloto Renardus. Os portais de assunto que, em Maio de 2001, participavam no projecto, eram os seguintes: DutchESS (*Dutch Electronic Subject Service*); NOVAGate (*Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences*); EELS (*Engineering Electronic Library, Sweden*); DAINet (*Deutsches Agrarinformationetz*); FVL (*The Finnish Virtual Library – Virtuaalikirjasto*); RDN (*Resource Discovery Network* com z.B.SOSIG, OMNI, Humbul, etc.); SSG-FI (*SonderSammelGebiets-FachInformationsführer, Vlib History Guide, Vlib Anglistik Guide, MathGuide, Geo-Guide*); Archivserver DEPOSIT.DDB.DE. Apresentavam-se ainda como futuros parceiros o DEF

elementos acordados de início, a solução passava por se adaptarem à nova situação ou aceitarem o facto de os seus registos não poderem ser incluídos em todas as opções de pesquisa; por outro lado, tornava-se necessário um acordo a nível semântico relativamente aos elementos de metadados e às regras de catalogação aplicáveis, incluindo o uso de vocabulários e notações normalizados.

Para que um portal de assunto possa considerar-se de qualidade controlada deverá responder a determinados requisitos, dos quais se destacam a divulgação da política de selecção e de desenvolvimento da colecção, a sua gestão (manutenção, verificação regular de conteúdos e URL e actualização permanente), a criação de metadados por especialistas de assunto e/ou bibliotecários, a descrição coerente, com base em metadados estruturados em campos semânticos e sintácticos bem definidos e baseados em normas internacionais, a indexação de todos os recursos de acordo com um esquema de classificação ou com outra linguagem documental controlada, opções de pesquisa por campos, através de operadores booleanos, pesquisa combinada usando vários elementos de metadados, ou por navegação numa árvore de assuntos, a utilização de normas que permitam a interoperabilidade com outros serviços da Internet (como o uso do conjunto de metadados Dublin Core, ou do protocolo Z39.50); deverá ainda incluir algumas mais-valias, como ficheiros de ajuda e informação sobre a dimensão e taxa de crescimento da colecção, a cobertura geográfica e linguística, os grupos-alvo do serviço, informação sobre o formato de metadados, com descrição semântica e sintáctica de cada elemento, sobre os sistemas de classificação e tesauros, entre outras (Neuroth e Koch, 2001, p.3-4).

Os parceiros do Projecto RENARDUS haviam já participado noutros projectos relacionados com os portais de assunto (como, por exemplo, o já referido projecto DESIRE e os projectos ROADS¹⁰⁷ e EELS¹⁰⁸) e tinham plena

fagportal (Danish Electronic Research Library), *Les Signets de la Bibliothèque Nationale de France* e BIBSYS.

¹⁰⁷ Resource Organisation And Discovery in Subject-based services. [Em linha]. [Consult. 15 Mar. 2005]. Disponível na WWW: <URL: <http://www.ukoln.ac.uk/metadata/roads/>>

¹⁰⁸ *Engineering E-Library, Sweden*. [Em linha]. [Consult. 15 Mar. 2005]. Disponível na WWW: <URL: <http://eels.lub.lu.se/>>

consciência das implicações que uma tal iniciativa traria a nível organizacional, já que “as barreiras técnicas à interoperabilidade são excedidas pelos desafios dos modelos organizacionais e de negócio usados” (Heerey, Carpenter, Day, 2001, p.1). Com efeito, durante o período em que se desenvolveu o projecto ocorreram várias transformações, quer a nível das ambições de serviço dos próprios portais, quer a nível das estruturas organizacionais que os forneciam. Esta ausência de estabilidade, reconhecida como inerente ao ambiente de serviços da Internet, constitui sempre um desafio às actividades que requerem um planeamento a longo ou a médio prazo (no caso em análise, um período superior a dois anos). No entanto, o Projecto procurou estabelecer uma estrutura cooperativa para os portais de assunto europeus, visando beneficiar simultaneamente os utilizadores e os próprios portais, na medida em que estes tirariam vantagens directas da partilha de soluções, fosse a nível de metadados, de acordos sobre normalização ou de soluções técnicas conjuntas.

A arquitectura do projecto RENARDUS consistia numa série de bases de dados interoperativas, pesquisáveis em simultâneo na Internet e cujos dados seriam acedidos pelo utilizador final através de uma interface *web*, como se pode ver no anexo 5. Cada participante ou grupo de participantes no Projecto deveria manter um servidor Renardus para a base de dados de conteúdos e para a informação administrativa¹⁰⁹. Os parceiros extrairiam os campos relevantes das suas bases de dados e exportá-los-iam, normalizados, de acordo com um modelo de dados pré-definido, para a base de dados Renardus local. Da informação administrativa constariam as descrições referentes à gestão da colecção e os mapas (esquemas resultantes dos mapeamentos). Tratando-se de um modelo distribuído, cada participante ficaria responsável pelo conjunto de informações existentes no seu servidor local Renardus, ou seja, pela produção, manutenção e qualidade dos seus serviços individuais. O protótipo do serviço Renardus consistiria num portal Z39.50 para a *web*, funcionando em conjunto com várias bases de dados Renardus distribuídas

¹⁰⁹ Cada serviço participante no Projecto RENARDUS teria de escolher um modelo para o servidor Renardus Z39.50 (ver anexo 5) e, através dele, implementar o acesso aos seus dados. O perfil Renardus estava de acordo, de um modo geral, com o *Bath Profile*, embora, de certa forma, o ampliasse, com o uso de campos como a descrição, a língua, o tipo e o país.

Z39.50, através do qual o utilizador final poderia efectuar as pesquisas e aceder aos seus resultados. Um portal deste tipo resultaria de compromissos entre o protocolo Z39.50 e o protocolo HTTP, já que, como atrás referimos, os dois não eram exactamente compatíveis. O Z39.50, baseado num modelo cliente-servidor, apresentava à partida as vantagens inerentes a uma norma, para além de ser um protocolo através do qual muitos dados estavam já acessíveis (nomeadamente catálogos bibliográficos nacionais). Para além disso adaptava-se a futuros desenvolvimentos, permitindo, pela sua característica modular, acrescentar sintaxes de interrogação, formatos de registos, semântica de campos e operações várias e, tratando-se de um protocolo de comunicação, não colocava grandes restrições às bases de dados que lhe estavam subjacentes. A sua funcionalidade e adequação aos objectivos do projecto Renardus são claras: quando um utilizador pretende interagir com a base de dados, o cliente (o programa usado para comunicar com a base de dados) abre uma sessão com o servidor e começa a enviar-lhe pedidos; este, por seu turno, envia-lhe as respostas. As operações de pesquisa e de recuperação estão separadas: uma pesquisa, que pode ser feita simultaneamente a várias bases de dados com uma única interrogação e utilizando várias sintaxes, tem como resultado um conjunto de registos, no servidor, que respondem à pergunta formulada; a totalidade dos conjuntos ou subconjuntos de registos pode ser recuperada, através da operação de recuperação, em diversos formatos, incluindo as variantes MARC, SGML/XML e texto livre. Além destas operações, o Z39.50 suporta ainda a função de controlo de acesso e a possibilidade de “*browsing*”.

A combinação das possibilidades de pesquisa e recuperação em sistemas heterogéneos oferecidas pelo Z39.50 com a facilidade de uso da *Web* era também um desiderato do projecto. Um portal que assumisse uma interface entre os protocolos Z39.50 e HTTP perderia em funcionalidade e flexibilidade em relação aos que possuíssem um cliente Z39.50 dedicado; contudo, os ganhos em facilidade de uso e acessibilidade poderiam ajudar a ultrapassar esse inconveniente (Falcoz, 2000). Segundo uma das publicações distribuídas durante o projecto RENARDUS (a D2.1, sobre as normas técnicas), a funcionalidade de um Z39.50 para portal *Web* podia dividir-se em três grupos

distintos: o conjunto de serviços Z39.50 suportados, o conjunto de facilidades de comunicação e o conjunto de serviços de portal¹¹⁰. A combinação da norma Z39.50 com a *Web* não só permitia interfaces simples para o utilizador, que rapidamente se poderiam desenvolver e adaptar, como possibilitava a existência de diferentes interfaces, de acordo com as preferências dos participantes no Projecto. Apontavam-se também as limitações da *Web* a nível da actualização dinâmica da interface do utilizador, por exemplo quando este pesquisava múltiplas bases de dados e pretendia ver os resultados assim que eles surgissem de uma das bases, recebendo depois as respostas das bases mais lentas. Como solução para ajudar a ultrapassar o problema indicava-se o uso da tecnologia *Java* e de um segundo canal de comunicação entre o cliente e o portal, embora se reconhecesse que algo se perderia em termos de simplicidade e facilidade de desdobramento do serviço daí resultante. Embora não existindo normas para a interface entre o Z39.50 e a *Web*, identificavam-se algumas formas normalizadas utilizadas por certos portais para proceder à interface de aplicações com um servidor *Web*¹¹¹. Especificava-se, pois, para o projecto, a norma Z39.50 como protocolo de pesquisa e a WWW como interface do utilizador. A XML (*eXtensible Markup Language*) era encarada com um certo cepticismo, já que não se reconhecia vantagem em converter os registos primeiro nesta linguagem antes de os normalizar em algo que o Z.39.50 pudesse entender (Miettinen, 2000); admitia-se, contudo, que ela poderia vir, no futuro, a beneficiar o projecto, numa fase de maior integração entre o Z39.50 e o RDF/XML e no processo de classificação automática e descrição de conteúdos dos recursos *web*.

Dada a grande diversidade existente entre os sistemas de classificação usados nos portais participantes (classificações universais, especializadas e mesmo de concepção local), todos eles foram mapeados num esquema comum de classificação, de modo a garantir a desejada consistência na pesquisa dos assuntos. O esquema adoptado foi a Classificação Decimal de Dewey (CDD),

¹¹⁰ Formatação e adaptação da formulação de uma pesquisa ("*query*") às bases de dados, incluindo o processamento de HTML, as formas de apresentação dos registos, incluindo a sua formatação para vários formatos de saída - apresentação, impressão ou descarregamento -, as facilidades ligadas à pesquisa paralela e ainda funcionalidades extra, como o armazenamento privado das formulações de pesquisa ("*queries*") e dos registos ou os alertas, entre outras.

¹¹¹ CGI, *Fast CGI* e os módulos *Apache*.

que se constituiu assim na linguagem de comutação (“*switching language*”) usada no Projecto Renardus. Do ponto de vista dos parceiros do projecto, as vantagens acrescidas que esta classificação apresentava, quando comparada com outras, eram: a sua disponibilidade em linha (*Web Dewey*), a adequação da sua estrutura e dimensão ao trabalho a desenvolver, a sua vasta utilização como linguagem de representação de conteúdos temáticos, o elevado número de recursos digitais já classificados de acordo com o seu esquema e a rapidez e frequência das suas utilizações. Considerava-se ainda vantajosa a possibilidade de otimizar a investigação e os desenvolvimentos metodológicos contínuos empreendidos pela OCLC¹¹², para além do facto de a CDD possuir já linguagens terminológicas mapeadas, como os LCSH¹¹³.

De acordo com as especificações constantes do documento distribuído em Julho de 2001 pelos responsáveis do Projecto¹¹⁴, o mapeamento das classificações deveria ser realizado por cada portal participante, que estabeleceria as equivalências entre a classificação usada localmente e a CDD. Cada mapa seria armazenado num simples formato ASCII, como¹¹⁵:

LN : Notação da classificação local
LC : Legenda (“*caption*”) da classificação local
RL : Código de relação
DN : Notação CDD
DC : Legenda (“*caption*”) CDD
UR : URL local de pesquisa para a classe
NT : Notas
: RE : Fim do registo

¹¹² Como atrás referimos, a OCLC Forest Press é a proprietária do esquema de classificação Dewey. O uso da CDD no projecto baseou-se num acordo de investigação com a OCLC Forest Press.

¹¹³ A OCLC iniciou, em 1994, a ligação dos LCSH à CDD (Vizine-Goetz *et al.*, 2004).

¹¹⁴ ARDÖ, Anders [et al.] – *Specification of Renardus Pilot broker system : Renardus deliverable D2.3 version 1.3*. 2001. [Em linha]. [Consult. 14 Out. 2002]. Disponível na WWW: <URL: <http://www.renardus.org/>>

¹¹⁵ Mapas traduzidos a partir de Ardö (2001, p. [25]).

Exemplo de um mapa da classificação Ei¹¹⁶:

LN : 441.1

LC : Barragens

RL : >

DN : 627.8

DC : Barragens e reservatórios

UR : <http://eels.lub.lu.se/ei/441.1.html>

NT : *alt.* : mapa para a maioria das subclasses de 627.8

: RE :

Para garantir a consistência em todo o processo, as relações do mapeamento foram estabelecidas entre pares de classes e não entre uma classe Dewey e os recursos individuais. O mapeamento fez-se numa única direcção, ou seja, da CDD para a classificação local e os portais foram aconselhados a terminar o mapeamento nos níveis hierarquicamente superiores (as notações de topo da hierarquia da classificação) antes de partirem, progressivamente, para os níveis mais baixos. Embora o objectivo final fosse o mapeamento entre todas as classes locais e a CDD, considerou-se prioritário incidir, em primeiro lugar, nas classes mais frequentemente utilizadas por cada portal local.

Os mapeamentos a realizar no âmbito do projecto RENARDUS pretendiam ir mais além da simples conversão entre dois sistemas de classificação, comum em projectos de mapeamento para uso em OPACs ou em catálogos colectivos, que estabeleciam simples ligações entre pares de classes sem especificação do grau e do carácter da equivalência existente entre elas. Pretendia-se que os mapeamentos RENARDUS possibilitassem aos utilizadores perceber que certas ligações de uma determinada classe da CDD podiam apontar para uma classe local de um portal com um conteúdo mais lato ou mais restrito, com maior ou menor sobreposição com a classe da CDD. Com esse objectivo, foram definidas 5 relações distintas de mapeamento:

¹¹⁶ Ei – *Engineering Index Thesaurus*, utilizado, entre outros, pelo portal EELS – Engineering E-Library, Sweden.

1. classe local totalmente equivalente à classe da CDD: o conteúdo temático da página local correspondia, de um modo geral, ao do assunto indicado na página Renardus;
2. classe local equivalente, de uma forma mais genérica, à classe da CDD: o conteúdo temático da página Renardus correspondia a um subconjunto do indicado na página local;
3. classe local equivalente, de uma forma mais específica, à classe da CDD: o conteúdo temático da página local correspondia a um subconjunto do indicado na página Renardus;
4. classe local com uma sobreposição máxima à classe da CDD: quando o conteúdo da página local representava uma grande parte da página Renardus e ainda outros assuntos relacionados;
5. classe local com uma sobreposição mínima à classe da CDD: existia alguma equivalência entre parte da página local e a página Renardus, mas também incluía outros assuntos relacionados.

Os três tipos de equivalência (mais genérica, mais específica e total) requeriam que uma das duas classes fosse um verdadeiro subconjunto da outra, excluindo a hipótese de poder ser mapeada a outra parte do esquema de classificação, ou que fossem ambas, basicamente, 100% equivalentes. As duas relações de sobreposição (máxima e mínima) pressupunham que partes de ambas as classes não pertencessem, claramente, ao conteúdo temático da outra classe.

Para realizar os mapeamentos adaptou-se uma ferramenta informática¹¹⁷ desenvolvida pelo projecto alemão CARMEN¹¹⁸. Os sistemas de classificação e a informação do mapeamento estavam alojados em diferentes servidores e cada portal participante enviava uma versão do seu esquema de classificação em formato legível por computador, para ser usado pela ferramenta de mapeamento. A interface para os utilizadores (portais participantes) consistia

¹¹⁷ A ferramenta Renardus, baseada na *web*, requeria o *software* de base de dados *mySQL*, um servidor *Web Apache*, *Javascripts* e *scripts PHP* do lado do servidor.

¹¹⁸ *Context Analysis, Retrieval and Metadata : Effective Networking*, projecto desenvolvido entre Outubro de 1999 e Fevereiro de 2002. Teve como foco a análise de conteúdo, visando atingir consistência e homogeneidade em ambientes descentralizados. <URL: <http://www.mathematik.uni-osnabrueck.de/projects/carmen/index.en.shtml>>.

em três janelas principais: uma para a notação local, outra para apresentar e percorrer a classificação-fonte (CDD) e uma terceira para receber e apresentar a informação do mapeamento, incluindo relações e notas. As relações eram apresentadas como ligações nas janelas das classificações, como se pode verificar no anexo 6.

Na síntese que apresentaram ao encontro satélite da IFLA em 2001¹¹⁹, os representantes do projecto levantavam algumas questões problemáticas ainda não resolvidas, como o grau de profundidade do mapeamento tanto a nível das classificações locais como da CDD, ou a exclusão dos auxiliares. O método seguido no mapeamento também apresentava algumas dificuldades, quando, por exemplo, uma área temática que constituía o foco de um portal correspondia a um nível profundo da CDD. Não estava também clara a forma de resolver o conflito por vezes existente entre as classificações especializadas numa determinada disciplina e a dispersão desta nos sistemas universais (como acontece, por exemplo, com a classificação especializada *Ei*, que possui 800 classes para a engenharia, disciplina que, na CDD, se encontra dispersa por 2300 categorias). Outras questões problemáticas apontadas diziam respeito ao grau de sobreposição de assuntos nos mapeamentos levados a efeito pelos parceiros e às inconsistências resultantes do uso, por parte dos portais, de mais do que um esquema de classificação.

Através do serviço Renardus, o utilizador final poderia navegar pela estrutura hierárquica da CDD (encarando-se a hipótese de eliminar, para este efeito, os ramos vazios dos níveis mais baixos da hierarquia) ou optar por fazê-lo através de uma estrutura gráfica, disponível em cada página de “*browsing*”, que ofereceria uma visão de todas as categorias que rodeavam o termo escolhido (a estrutura gráfica de navegação, tomando como exemplo a pesquisa do assunto “Arquitectura”, pode ver-se no anexo 7). A pesquisa avançada utilizaria as informações constantes no mapeamento para tornar possível a pesquisa de toda a informação de assunto existente nos portais a nível local (palavras-

¹¹⁹ KOCH, Traugott; NEUROTH, Heike; DAY, Michael (2001) – Renardus : cross- browsing European subject gateways via a common classification system (DDC). In IFLA SATTELITE MEETING, Dublin (OH), 2001 - *Subject retrieval in a networked environment : proceedings..* München : K.G. Saur, 2003. p. 25-33.

chave não controladas, descritores e cabeçalhos de assunto, notações), através do elemento genérico, e das classes CDD mapeadas, através do elemento de classificação CDD. O utilizador poderia seleccionar as entradas de assunto encontradas numa pesquisa, fazendo-o numa janela de índice que aquela, normalmente, gera.

Realizado o trabalho, no âmbito do Projecto, pelos portais participantes, desejava-se que, no futuro, os proprietários dos sistemas de classificação (ou outros utilizadores) mantivessem um mapeamento com a CDD – para o qual seria necessário obter licenças de utilização das classificações. Considerava-se ainda imprescindível a normalização desse trabalho através do uso de esquemas RDF para os vocabulários envolvidos, para as suas relações semânticas e para os diferentes tipos de mapeamento. Visava-se ainda elaborar recomendações relativas ao acesso por assunto em portais e serviços intermediários (“*broker services*”), incluindo aspectos como a consistência, a precisão e a facilidade de utilização do mapeamento.

Com o objectivo de conduzir o Renardus a um estatuto de serviço, promovendo o seu desenvolvimento e o alargamento a outros membros, criou-se, em Setembro de 2002, o *Renardus Consortium*. Um Grupo de Gestão é responsável pelas negociações com potenciais parceiros e patrocinadores, pela manutenção do serviço, pela comunicação e pela difusão do mesmo¹²⁰. Foram igualmente elaboradas e divulgadas directrizes que incluem os requisitos organizacionais, técnicos e de metadados a ser tidos em consideração pelos potenciais portais participantes no serviço¹²¹.

¹²⁰ Para informações sobre o Consórcio e o seu Grupo de Gestão veja-se <URL: http://www.renardus.org/about_us/organisation.html>. O modelo organizacional para o serviço intermediário RENARDUS, apresentado no final do projecto, consta do anexo 8.

¹²¹ Disponíveis em http://www.renardus.org/about_us/guidelines/

5.2 MACS (Multilingual Access to Subjects) – um exemplo de interoperabilidade entre linguagens de indexação de diferentes bibliotecas nacionais, em diferentes línguas

O projecto MACS (*Multilingual Access to Subjects*) surgiu em 1997, como resposta à vontade expressa pela CENL (*Conference of European National Librarians*) de resolver o problema do acesso por assunto nas bases de dados bibliográficos europeias. A barreira da língua no acesso aos catálogos fazia-se sentir em países de natureza multilíngue como, por exemplo, a Suíça, e a expansão da Internet veio despoletar igualmente noutros a necessidade de transpor o obstáculo linguístico a fim de facilitar um acesso internacional aos seus recursos. Iniciado sob os auspícios do Grupo de Trabalho sobre o Acesso por Assunto Multilíngue do Fórum CoBRA+¹²², o projecto MACS desenvolveu-se através da cooperação de quatro bibliotecas nacionais - a Biblioteca Nacional Suíça (SNL), a Biblioteca Nacional de França (BnF), *Die Deutsche Bibliothek* (DDB) e a *British Library* (BL) – que aceitaram definir uma ideia-base para o acesso por assunto multilíngue e demonstrar a sua viabilidade através de um estudo que decorreu entre o Outono de 1997 e Fevereiro de 1999.

Todas as bibliotecas envolvidas neste projecto utilizavam linguagens de indexação normalizadas, monolíngues: a SWD / RSWK (*Schlagwortnormdatei / Regeln für den Schlagwortkatalog*), em língua alemã, usada pela Biblioteca Nacional da Suíça e pela Biblioteca Alemã (*Die Deutsche Bibliothek*), o RAMEAU (*Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié*), usado na Biblioteca Nacional de França e os LCSH (*Library of Congress Subject Headings*), utilizado na *British Library*. O projecto MACS definiu como objectivo fornecer os instrumentos através dos quais se pudesse

¹²² CoBRA+: acção concertada de bibliotecas nacionais e agências bibliográficas da Europa, iniciada em 1996 com financiamento da Comissão Europeia (DGXIII/E) que deu continuidade aos trabalhos de investigação e desenvolvimento realizados no âmbito de CoBRA (*Computerized Bibliographic Record Actions*) com o objectivo de promover os serviços bibliográficos internacionais fornecidos pelas agências nacionais e a coerência entre as iniciativas nacionais numa perspectiva comunitária, alargando-se também às questões referentes ao desenvolvimento das publicações electrónicas e dos serviços com elas relacionados. O enfoque CoBRA+ é colocado sobretudo no acesso e partilha de recursos, nas necessidades do utilizador e no uso das redes de comunicações (Cf. *Computerized Bibliographic Record Actions* [Em linha]. [Consult. 27 Jun. 2003]. Disponível na WWW: <URL: http://www.kb.nl/gabriel/projects/en/partnerships_frame_en.html>).

“aceder às bases de dados das bibliotecas numa base multilíngue graças ao uso de cabeçalhos equivalentes provenientes de *ficheiros de autoridade de assunto*”¹²³, utilizando para tal, por um lado, “o ambiente estável do *controlo de autoridade*” e, por outro, o processo de ligação baseado na tecnologia Internet “para estender o uso das linguagens de indexação por assunto a um nível multilíngue conectado a uma escala internacional” (Clavel-Merrin, 2003a, p.1). Recorreu-se, assim, a três linguagens solidamente estruturadas a nível monolíngue, com controlo de autoridade assegurado e com uma utilização bastante alargada¹²⁴, garantindo-se, por um lado, uma maior eficácia no estabelecimento de equivalências entre elas e permitindo, por outro, um rápido alargamento do acesso multilíngue a uma série de outras bases de dados utilizadoras das mesmas linguagens e, por isso, em condições de poderem usufruir, muito rapidamente, dos resultados do projecto.

O início do projecto MACS coincidiu com a reformulação do catálogo de assuntos da Biblioteca Nacional Suíça, que tinha optado pela utilização dos SWD/RSWK para a indexação das suas colecções, pondo fim ao catálogo de assuntos baseado na CDU – Classificação Decimal Universal. Embora se reconhecesse que o acesso por assunto se deveria fazer em alemão, francês, inglês e italiano, a escolha dos SWD teve em consideração o facto de a maioria dos documentos tratados (ca 60%) serem em língua alemã, assim como ser esta a língua principal de trabalho na biblioteca, para além do facto de 75% da população suíça ter como língua materna o alemão (Landry, 1997, p. 3). Iniciava-se assim uma colaboração muito estreita com a *Deutsche Bibliothek*, o que só viria facilitar o trabalho encetado no âmbito do projecto MACS.

Partindo do princípio de que cada língua é o reflexo de uma cultura e que, conseqüentemente, linguagens de indexação em diversas línguas reflectem também culturas diferentes, o Projecto MACS apresentou como propósito o

¹²³ Cit. por Clavel-Merrin, 2003a, p. 2.

¹²⁴ O SWD/RSWK, sistema criado no início da década de 1980, para além de largamente utilizado na Alemanha é também usado na Áustria, os LCSH nos países de língua inglesa e em instituições internacionais e o RAMEAU, para além de ser utilizado em França, é-o igualmente em alguns países francófonos. O RAMEAU foi constituído a partir do *Répertoire de vedettes-matière* (RVM) da Biblioteca da Universidade de Laval (Québec, Canadá), adaptação em língua francesa dos LCSH (Cf. Landry, 1997, p. 6).

estabelecimento de equivalências entre linguagens já existentes nas três línguas seleccionadas (e não a criação de um novo tesouro multilingue ou a simples tradução de termos de indexação), com o objectivo de permitir o acesso, com uma única chave de pesquisa, à informação indexada por diferentes bibliotecas e em diferentes idiomas, sendo a língua utilizada na pesquisa independente da língua usada na indexação - podendo utilizar-se, por exemplo, um termo em alemão para pesquisar a informação indexada com os termos equivalentes em inglês e em francês. Estabelecia-se, assim, uma clara distinção entre a língua do documento, a língua do catálogo (usada pelas linguagens de indexação) e a língua de pesquisa (usada pelo utilizador). Tal como aconteceu nos trabalhos do projecto RENARDUS (ainda que com linguagens controladas diferentes), também aqui o mapeamento dos termos representativos dos conceitos nas três linguagens seleccionadas pelo Projecto exigiram um trabalho intelectual para identificação das equivalências exactas ou parciais, simples ou complexas. Foram também estudados diferentes modelos de ligação que pudessem ser usados pelo projecto (estes são apresentados no anexo 9). O sistema completo MACS, cuja arquitectura consta do anexo 10, apresentaria várias bases de dados de cabeçalhos de assunto distintas, interligadas, cada uma delas mantida pelos respectivos proprietários. Estas encontrar-se-iam na periferia do sistema. No núcleo central estaria a base de dados de ligação MACS, contendo dados adicionais, que podiam ou não ser usados pelos parceiros, para sugerir ligações entre os vários cabeçalhos de assunto das bases de dados individuais. Esta base central conteria ligações (agrupamentos - "*clusters*" - de expressões em várias linguagens de assunto, cada uma das quais contendo termos e operadores para os ligar, geralmente o operador "E", embora a maior parte das expressões consistisse num único termo).

Após o estudo de viabilidade do projecto, que decorreu entre 1997 e 1999 e para o qual se seleccionaram, para efeitos do mapeamento, as áreas temáticas "Teatro" e "Desporto", desenvolveu-se um protótipo¹²⁵ – aplicação *web* baseada em normas existentes (como os formatos MARC e a norma Z39.50) e

¹²⁵ O protótipo foi desenvolvido pela Index Data (Dinamarca) e pela Biblioteca da Universidade de Tilburg (Holanda).

independente dos sistemas dos parceiros envolvidos – que permitiu desenvolver os mecanismos para gerir equivalências entre cabeçalhos, incluindo ligações múltiplas, já que um termo numa linguagem pode ser equivalente a mais do que um termo noutra linguagem (situação em que se utilizam os operadores booleanos) – exemplos de mapeamento entre as três linguagens podem ser vistos no anexo 11. Para além de possuir uma interface trilingue, a aplicação teve de obedecer a imperativos de flexibilidade e exequibilidade em termos financeiros, a fim de permitir a sua operacionalidade e a futura integração de novos catálogos e novas linguagens. No protótipo, a pesquisa é feita numa língua escolhida pelo utilizador (que corresponde a uma das linguagens de indexação adoptadas) e é lançada nos catálogos locais, usando o protocolo Z39.50 e as ligações de equivalência estabelecidas na interface de gestão de ligações. Podem percorrer-se todos os catálogos, ou seleccionar apenas alguns - no anexo 12 é possível seguir os vários passos de uma pesquisa no protótipo MACS. Na opção “*browse translations*” tem-se acesso ao mapa de equivalências de um determinado assunto, nos seus vários aspectos ou facetas, entre os LCSH, o SWD e o RAMEAU. Uma vez que o projecto utiliza o ambiente estável do controlo de autoridade (Clavel-Merrin, 2003), o mapeamento completo inclui, não só os termos equivalentes nas três línguas, mas também, para cada um deles, o número de identificação na linguagem de indexação respectiva.

De acordo com Freyre e Naudi (2003, p.7), o MACS funciona como uma ponte entre sistemas concebidos para organizar e designar conceitos, sendo, de facto, um dicionário de ontologias e podendo tornar-se uma via para a construção de tesouros multilingues a partir de sistemas de indexação já existentes. Num estágio mais avançado de cooperação o MACS pode ser usado no processo de indexação das bibliotecas participantes, uma vez que permite reutilizar a indexação já realizada numa língua estrangeira ou mesmo “traduzir” a indexação realizada na língua do próprio documento, de modo a que a indexação original seja feita uma única vez - facilitando assim o tratamento dos conteúdos a indexadores que, por exemplo, não dominem a língua em que os documentos estão escritos.

O modelo organizacional definido pelos parceiros do projecto, de acordo com o qual se gerem as ligações¹²⁶, permite uma gestão flexível e partilhada das relações entre as diferentes linguagens de indexação, tendo em linha de conta, por um lado, a automatização de certos processos e, por outro, o trabalho a desenvolver com as redes nacionais para distribuir e tornar mais rápido o trabalho a realizar. Obedece, em linhas gerais, aos mesmos princípios de cooperação subjacentes ao projecto, que são, genericamente, a autonomia de cada linguagem de indexação, a responsabilidade limitada de cada parceiro no estabelecimento de ligações, a capacidade para propor e o direito de comentar o trabalho realizado pelos outros parceiros, sempre numa relação de respeito mútuo pelas propostas e soluções encontradas. As equivalências entre os termos são estabelecidas por especialistas de assunto. Para o protótipo, seleccionou-se uma área temática em relação à qual se compararam as respectivas listas monolingues de termos de indexação. Analisaram-se as ligações a nível terminológico (termo de indexação, cabeçalho de assunto), a nível semântico (registo de autoridade) e a nível sintáctico (aplicação), trabalho para o qual as equivalências sistematicamente estabelecidas pela BnF entre a linguagem RAMEAU e os LCSH, a nível dos seus registos de autoridade, se revelaram de grande utilidade¹²⁷. Decidiu-se, assim, proceder ao carregamento no sistema dos campos adequados do registo RAMEAU (bem identificados no formato Unimarc), de modo a alimentá-lo automaticamente com um elevado número de ligações RAMEAU – LCSH. Tratando-se de um universo de aproximadamente 70 000 ligações, os parceiros do projecto reconheceram a necessidade de fasear o trabalho, executando-o por temas, de modo a permitir a confirmação de um correcto desenvolvimento em inglês (por parte da BL), a criação de equivalências em alemão (retiradas da SWD pela DDB e pela SNL) e uma avaliação final de todas as ligações estabelecidas em cada área temática. Apontaram ainda a necessidade de desenvolver novas especificações, para permitir os descarregamentos e a automatização das actualizações de termos de indexação e cabeçalhos já existentes no MACS, a

¹²⁶ Sistema disponível através da interface de gestão de ligações.

¹²⁷ Esta equivalência foi certamente muito facilitada pelo facto de o RAMEAU corresponder a uma adaptação, em língua francesa, dos LCSH, para cuja estrutura nem os indexadores francófonos do Canadá nem os indexadores e especialistas franceses terão contribuído (Holzem, 1999, p. 132-133).

partir dos sistemas locais. Foi igualmente decidido, pelos parceiros, acrescentar um campo de classificação à interface de gestão das ligações, aproveitando também o facto dessa informação existir já nos registos RAMEAU, correspondendo aos domínios da Classificação Decimal de Dewey. Embora não se tenha ainda chegado a um consenso sobre se será esse ou não o sistema de classificação a adoptar pelo MACS, a utilização de uma classificação permitirá criar grupos homogéneos de termos de indexação, facilitando a distribuição de trabalho pelos parceiros e pelas redes a que cada um deles se encontra ligado. Para iniciar o trabalho, foram seleccionadas duas áreas temáticas: Matemática e Psicologia.

A solução que o MACS propõe para o acesso aos assuntos procura utilizar as redes existentes a nível nacional e internacional, otimizar o trabalho também já existente a nível de linguagens de indexação monolíngues bem estruturadas e consolidadas pelo uso e desenvolver a cooperação internacional. Além disso, preconiza a expansão do sistema através da possibilidade de um futuro acesso directo às bases de autoridade locais, no sentido de melhorar a pesquisa (através das referências “Veja” e “Veja também”) e da integração do trabalho diário desenvolvido pelos parceiros. Por constituir um meio viável para assegurar a pesquisa cruzada entre linguagens em diferentes línguas e visando alargar-se a outras bibliotecas nacionais, propõe-se ainda vir a integrar o serviço TEL (*The European Library*)¹²⁸ como mais um ponto de acesso adicional (Landry, 2003).

5.3 O projecto HILT (High Level Thesaurus) – uma proposta de interoperabilidade entre diferentes linguagens de indexação utilizadas por entidades tipologicamente distintas (bibliotecas, arquivos, museus)

O projecto HILT (High Level Thesaurus) desenvolveu-se por duas fases, a primeira, com a duração de um ano, entre o último trimestre de 2000 e o final de Setembro de 2001¹²⁹ e a segunda entre Maio de 2002 e Setembro de 2003,

¹²⁸ <URL: <http://www.europeanlibrary.org>>.

¹²⁹ O relatório final data de Novembro de 2001.

datando o último relatório de avaliação externa desta fase do projecto de Dezembro de 2004.

O objectivo do HILT – Fase I, financiado pela RSLP (*Research Support Libraries Programme*) e pelo JISC (*Joint Information Systems Committee*), foi desenvolver um estudo sobre o problema da pesquisa cruzada e do “*browsing*” por assunto através de vários tipos de comunidades, serviços e recursos no Reino Unido, tendo em consideração a grande diversidade de sistemas de tratamento dos assuntos e de práticas associadas existentes e contemplando bibliotecas, museus, arquivos, instituições de ensino superior, de formação contínua, serviços electrónicos, catálogos distribuídos baseados no Z39.50, serviços e redes como o DNER (*Distributed National Electronic Resource*) e a RDN (*Resource Discovery Network*), bases de dados bibliográficos, entre outros (Nicholson e Wake, 2003, p. 61). O projecto reunia como parceiros o UKOLN (*UK Office for Library and Information Networking*), a MDA (*Museum Documentation Association*), o *National Council on Archives*, a OCLC (*Online Computer Library Center, Inc.*), o *Scottish Library and Information Council* e a SUfi (*Scottish University for Industry*) e foi conduzido pelo Centro de Investigação em Bibliotecas Digitais da Universidade de Strathclyde (Glasgow).

Antes mesmo do projecto ter início, decidiu-se, de acordo com as entidades financiadoras, que os primeiros passos a dar consistiriam: no levantamento e revisão tanto da literatura existente como da opinião de especialistas; na identificação de comunidades, serviços e iniciativas-chave, das suas perspectivas em relação ao tratamento dos assuntos e das exigências dos seus utilizadores; no reconhecimento dos diferentes sistemas em uso (e de outros susceptíveis de poderem vir a ser considerados como possíveis soluções); e na identificação de questões relevantes de carácter organizacional e inter-organizacional (como, por exemplo, custos de conversão retrospectiva de metadados). O acordo sobre um possível caminho a seguir basear-se-ia na análise e organização dos dados recolhidos, na elaboração de mapas e discussão dos seus resultados, bem como na produção de um relatório intermédio, tudo questões a debater em profundidade num *workshop* a realizar antes do final do projecto. Um questionário foi lançado a 46 depositários de

informação (a maioria dos quais do Reino Unido, mas também da Austrália, Canadá e Estados Unidos), tendo sido recebidas 42 respostas. O questionário estruturava-se em quatro secções: acesso aos assuntos, acesso aos dados, requisitos do pessoal e utilizadores, problemas e questões a debater. Apurou-se que 64% dos inquiridos que responderam ao inquérito usavam esquemas adaptados a uso local, a fim de ultrapassarem problemas como as lacunas (ou fraco tratamento) de algumas áreas temáticas, ou para contemplarem novos conceitos ainda não representados nas linguagens documentais, entre outras razões. A conversão retrospectiva, assim como os dados já "herdados", apresentava-se também como um problema de difícil resolução. Das respostas apurou-se ainda que 36% dos depositários de informação facilitavam a pesquisa cruzada através do protocolo Z39.50 e 17% através da *Web*, existindo também acessos via Telnet e motores de busca da Internet. 95% das respostas revelavam a necessidade expressa dos utilizadores em aceder por assunto aos recursos existentes no depositário e 83% consideravam importante poderem efectuar pesquisas cruzadas por assunto, acedendo também a outros catálogos; por parte do pessoal essa necessidade era igualmente manifesta (93% no primeiro caso e 79% no segundo). Os depositários de informação encaravam ainda, maioritariamente (62%), a possibilidade de adopção de um novo tesouro de alto nível¹³⁰, se ele fosse amplamente usado por outros serviços e de interesse para os utilizadores; 36% faziam depender a sua decisão da ponderação de vários factores, como a medida em que o novo esquema beneficiaria o depositário, a comunidade em que estava inserido e o utilizador, se trazia ou não benefícios adicionais em relação ao esquema existente, qual o grau de diferença entre ambos, se a mudança se encaixava nos objectivos organizacionais, se era viável e valia a pena o tempo, esforço e dinheiro despendidos para empreender a mudança ou se havia ou não possibilidade de os dados referentes aos assuntos já existentes serem mapeados com o esquema proposto (Wake e Nicholson, 2001).

¹³⁰ Por tesouro de alto nível entende-se um tesouro que contém os termos mais elevados de uma hierarquia, ou os termos mais genéricos.

O *workshop* teve lugar em Junho de 2001 e reuniu um grupo de 50 delegados oriundos de arquivos, bibliotecas, museus e serviços de informação em linha. Para além da apresentação das diferentes perspectivas das comunidades envolvidas, foram analisadas e discutidas questões como o mapeamento de terminologias, interfaces de tesouros e futuros desafios da *web* semântica. Após o levantamento efectuado sobre os esquemas usados pelos depositários de informação seleccionados¹³¹, concluiu-se que, na maioria das comunidades inquiridas, existiam um ou dois esquemas dominantes, ou a combinação de esquemas: nos arquivos, por exemplo, era o tesouro da UNESCO o predominante, enquanto nas bibliotecas e serviços de informação em linha predominava a CDD combinada com os LCSH; a comunidade dos museus, embora usando também os LCSH, revelou um considerável investimento na construção de vocabulários controlados vocacionados especificamente para objectos, em determinadas áreas temáticas. Denotou-se ainda, em todas as comunidades, a existência de esquemas próprios. O facto de muitas das entidades depositárias não aplicarem a sua forma de abordagem preferida, de modo consistente, em todos os seus recursos, intensificava o problema da pesquisa cruzada e foi reconhecido que tal se devia, fundamentalmente, a duas razões: a impossibilidade de proceder à conversão retrospectiva dos registos quando ocorriam actualizações aos esquemas usados, dada a dimensão das colecções e a falta de meios; e a adição de novos termos, por parte dos depositários, quando a linguagem escolhida não possuía os adequados para representar determinados assuntos. A análise e representação de assuntos das várias comunidades-alvo foram também caracterizadas da seguinte forma:

I. Arquivos

Nos arquivos, a indexação das colecções era recente e não universal, representando a indexação a nível do item um trabalho imenso; a abordagem arquivística tornava a atribuição de "assunto" problemática, havendo divergência em relação às normas das bibliotecas; quanto aos instrumentos

¹³¹ Os depositários de informação foram seleccionados entre as comunidades do Reino Unido identificadas como alvo do projecto: arquivos, bibliotecas, museus, serviços de informação em linha, iniciativas governamentais. Veja-se o quadro resultante do levantamento no anexo 13.

utilizados para indexar, salientava-se o uso crescente do tesouro da UNESCO e também o uso substancial dos LCSH.

II. Museus

Nos museus a indexação era historicamente encarada como pouco importante e, quando realizada, era-o muitas vezes sem um vocabulário controlado ou com um vocabulário controlado específico da instituição, já que se consideravam os esquemas desenvolvidos para as bibliotecas inadequados aos objectos de museu; o “assunto” podia depender da influência ideológica ou temática do museu ou da colecção, denotava-se uma certa resistência a tornar disponível a informação e em aderir a normas e a existência de uma quantidade imensa de objectos não indexados; o AAT (*Art and Architecture Thesaurus*) vinha a ser mais aceite e para áreas mais vastas do que a sua designação implicava; a SHIC (*Social History and Industrial Classification*) era usada mas não actualizada, existindo ainda uma lista de nomes de objectos mantida pela MDA.

III. Bibliotecas

As bibliotecas possuíam grandes quantidades de material legado, frequentemente com utilização de diferentes versões de esquemas num mesmo catálogo, existiam variações locais dos esquemas utilizados, sendo os mais comuns (mas ainda minoritários) os LCSH e a CDD; os esquemas britânicos / europeus revelavam-se mais relevantes mas os dos Estados Unidos eram melhor mantidos.

IV. Serviços de informação electrónica

Os serviços de informação electrónica desenvolviam os seus próprios esquemas para servir os seus utilizadores, uma vez que os existentes nem sempre se consideravam apropriados; os seus conteúdos eram cada vez mais catalogados pelos seus autores em vez de o serem por profissionais; sentia-se, nesta comunidade, a falta de um controlo centralizado para coordenar o uso de vocabulário normalizado.

As diferenças detectadas nestas quatro comunidades indiciavam alguma dificuldade em adoptar uma abordagem baseada num único esquema de assuntos e foi encarada a possibilidade de mapear ou comutar os esquemas já usados, para o que seria necessário investigar as terminologias do utilizador e as formas de os mapeamentos serem usados para criar ferramentas de pesquisa amigáveis que contemplassem as complexidades desses esquemas. Tornava-se igualmente necessário investigar a possibilidade de iniciar o trabalho inter-comunidades com a descrição a nível de colecção. Para melhor compreender a natureza dos problemas a resolver relativamente aos esquemas a adoptar ou mapear, procedeu-se à identificação dos pontos fortes e fracos dos esquemas mais significativos (AAT, CDD, LCSH, CDU e tesauro da UNESCO), cujo quadro-síntese consta do anexo 14.

Perante um vasto leque de opções colocadas durante o *workshop*, que podem ser consultadas no anexo 15, concluiu-se que o melhor caminho a seguir (provavelmente o único consensual) seria o do mapeamento, que apresentava também a vantagem de mais rapidamente poder apresentar resultados. Foi assim decidido avançar com um serviço piloto de mapeamento, combinado com a criação de uma *task force* ou “agência de terminologias” que facilitaria o envolvimento e o controlo, a um nível muito prático, por parte das comunidades envolvidas. Este serviço procederia ao mapeamento de esquemas-chave como os LCSH, o tesauro da UNESCO, a CDD, a CDU, o AAT e, talvez, terminologias regionais e de utilizadores e adaptações locais de esquemas *standard*, com a possibilidade de utilizar um deles, por exemplo a CDD, como “espinha dorsal” da abordagem. Os utilizadores lançariam as suas pesquisas através de uma terminologia que lhes fosse significativa, especificando, se necessário, o contexto, e obteriam uma lista de termos equivalentes ou quase-equivalentes com os quais poderiam realizar uma pesquisa cruzada ou efectuar *browsing* através dos vários serviços. Para tal reconhecia-se a necessidade de aprofundar os trabalhos sobre a interface de um tal sistema com muitos serviços, em particular os serviços Z39.50, mas considerava-se que o trabalho já desenvolvido pelo CAIRNS, possibilitando a interface de clientes Z39.50 baseados na *web* com bases de dados contendo descrição SQL a nível de colecção, demonstrava a sua viabilidade. Tratando-se de um serviço-piloto,

permitia ainda, através de uma recolha mais alargada de informação, aprofundar uma série de questões importantes para a análise custo-benefício (custos, necessidades dos utilizadores, escolha da terminologia central do sistema - a do utilizador ou a CDD -, a melhor forma de integrar os desenvolvimentos da *web* semântica e da inteligência artificial, ou o *design*, entre outras) antes de se assumir um compromisso a longo prazo. Os participantes no *workshop* consideraram que este serviço-piloto deveria estar fortemente centrado no utilizador, deveria obter dados realistas sobre custos e incluir uma análise custo-benefício em diferentes níveis do serviço e do mapeamento; deveria ainda procurar envolver, a nível do financiamento e gestão, agentes internacionais reconhecidos, tendo como objectivo uma rápida implementação; reconheciam a conveniência em abarcar um vasto leque de serviços-alvo e em utilizar, sempre que possível, mapeamentos legíveis por computador já existentes; defendiam a forte ligação a uma *task force* entre domínios e sectores, e apontavam a necessidade de definir “mapeamento” *task force*” e “agência de terminologias” de uma forma mais específica. Na documentação que acompanhou o *workshop*¹³² descrevia-se a hipótese de uma agência terminológica do Reino Unido que poderia ter sub-agências regionais e actuaria como uma superestrutura de arquivos, bibliotecas e serviços electrónicos, coordenando e integrando todos os esforços de investigação sobre as questões terminológicas, aconselhando sobre as melhores práticas e as melhores soluções financeiramente sustentáveis, estimulando a utilização e adesão a normas e outros *standards*, coordenando e mantendo, se necessário, um serviço de mapeamento, estabelecendo canais de comunicação com o objectivo de minimizar as variantes quando os esquemas fossem adaptados ou acrescentados, fomentando a discussão a nível internacional, assegurando-se da compatibilidade mútua entre os resultados atingidos no Reino Unido e os desenvolvimentos e acordos a nível internacional.

A Fase II do Projecto HILT, desenvolvida a partir dos resultados da Fase I e prevista para durar 12 meses, correspondeu ao desenvolvimento do projecto-piloto, focando-se nos requisitos de terminologia e tesouros ao nível da

¹³² Cf. *HILT workshop breakout sessions: discussion issues* (2001), p. 5.

colecção, embora tendo igualmente presente a necessidade de recuperação ao nível do item. Utilizando os trabalhos empreendidos durante a fase I, propunha-se elaborar um roteiro ("*route map*") para terminologias, o serviço TeRM, visando obter uma experiência prática de investigação e o estabelecimento de requisitos de um serviço de terminologias de assunto para o ambiente informativo JISC, com particular referência ao DNER e ao RDN, ao utilizador, ao nível de colecção, à compatibilidade internacional e às reflexões sobre um acesso amplo a nível local, regional e de todo o Reino Unido. Esta segunda fase visava ainda apresentar recomendações para a eventualidade de um futuro serviço, tendo em consideração factores como o nível e a natureza das necessidades dos utilizadores, a praticabilidade, os requisitos do *design*, a efectividade, a disponibilidade de pacotes de *software* já existentes e, sobretudo, a relação custo-benefício (respeitante aos utilizadores do ensino superior e da formação contínua) de um serviço global de terminologia focalizado, em primeiro lugar, nas necessidades a nível de colecção (Nicholson, 2002, p.2). Os participantes na Fase II do projecto foram basicamente os mesmos da Fase I, embora com um maior envolvimento do DNER, da RDN e das instituições de formação contínua, e integraram também esta fase do HILT os especialistas em terminologia Alan Gilchrist e Leonard Will, este como avaliador externo. O Centro de Investigação em Bibliotecas Digitais da Universidade de Strathclyde, líder do projecto, devido ao seu envolvimento em projectos anteriores¹³³ e na 1ª fase do HILT, "possuía uma larga experiência na utilização das descrições a nível de colecção num ambiente dinâmico distribuído e nos problemas terminológicos associados" (Nicholson, 2002, p.3), para além de disponibilizar um ambiente de informação distribuída bastante rico para "laboratório" do projecto-piloto, possibilitando ainda a sua interacção com utilizadores e serviços¹³⁴. A OCLC deu também o

¹³³ Participou nos projectos CAIRNS (que utilizou os pontos fortes da colecção para desenvolver mini-*clumps*), SCONE (*Scottish Collections Network*) e SEED, financiado pelo *Scottish Executive Education Department* (que se combinaram para construir uma base de dados de colecções inter-sectoriais).

¹³⁴ Incluem-se neste ambiente o catálogo distribuído das universidades CAIRNS (*Cooperative Information Retrieval Network for Scotland*), a Biblioteca Nacional da Escócia (NLS), a NGfL (*National Grid of Learning*), o SLAINTE (*Scottish Libraries Across the Internet*), as bases de dados da Biblioteca Digital de Glasgow (GDL) e de colecções SCONE, um servidor de impressos electrónicos OAI (*Open Archives Initiative*), a base de dados NOF (*New Opportunities Fund*) e outras, igualmente correspondentes a projectos de digitalização, além do potencial de instalar outras bases de dados Z39.50.

seu apoio, permitindo o acesso ao mapeamento legível por computador entre a CDD e os LCSH.

Por razões práticas e após análise dos produtos existentes no mercado, o projecto-piloto TeRM utilizou o *software* comercial *Wordmap*, embora sem a intenção de vir a influenciar ou favorecer de algum modo a sua utilização futura. A nível terminológico, baseou-se nas terminologias RDN, nas disponíveis no próprio *Wordmap* que incluíam, em especial, os termos geralmente usados pelos utilizadores da *Internet*, e em subconjuntos seleccionados dos LCSH, da CDD, do tesouro da UNESCO e do AAT, para além do mapeamento LCSH-CDD e da possibilidade de vir ainda a incluir os mapeamentos CDD – cabeçalhos de assunto *Conspectus*¹³⁵.

Retomando as questões levantadas no final da Fase I do Projecto, a primeira que se colocou foi em relação à necessidade ou não de o TeRM possuir uma estrutura central, em função da interacção a estabelecer com o utilizador: entre este e um único esquema central ao qual todos os outros se encontrassem mapeados no TeRM ou entre o utilizador e cada esquema individual. Na hipótese de a melhor escolha ser a primeira, a questão era saber qual o esquema central a adoptar, se existente – e o apontado como mais provável foi a CDD, por se tratar de uma linguagem bem estruturada, hierárquica, traduzida em muitas línguas. A segunda questão era saber se, a longo prazo, a melhor solução para o problema das terminologias de assunto num ambiente distribuído seria a identificação e adopção de um único esquema, aceite como adequado para cobrir todos os aspectos de todos os domínios, ou de um esquema totalmente novo ou já existente mas possivelmente alterado para responder aos requisitos de um modelo que fosse genericamente aceite.

A construção do ambiente de pesquisa foi feita juntando várias colecções DNER e RDN, colecções de arquivos, de museus, e uma colecção local OAI a uma cópia da base de dados SCONE. Criou-se assim uma base de dados de colecções para teste do HILT Fase II, e um ambiente de pesquisa cruzada

¹³⁵ Igualmente cedido pela OCLC. Os cabeçalhos de assunto *Conspectus* eram também usados na base de dados de colecções do CAIRNS.

baseado em CLD (*Collection Level Description*), usando o mecanismo dinâmico desenvolvido pelo CAIRNS. O objectivo era utilizar “esquemas de assunto nativos” e usar o TeRM como forma de retirar a ambiguidade aos termos do utilizador e resolver as diferenças existentes entre os esquemas - no anexo 16 podemos ver o diagrama de funcionamento do TeRM (*Interactive Terminologies Route Map*).

6. Uma proposta de desenvolvimento da PORBASE no contexto do acesso por assuntos à informação em rede

Como já atrás referimos, a PORBASE tem crescido, até ao presente, com base na constante integração de registos bibliográficos da comunidade cooperante. A base cresce assim quer em termos de diversidade (já que muitos dos registos provenientes das bibliotecas cooperantes não se encontram referenciados na base e não existem na Biblioteca Nacional), quer sobretudo em quantidade de existências que lhe são acrescentadas – é igualmente grande o número de registos que já existem na base e aos quais apenas se acrescenta uma nova cota, associada à sigla da biblioteca a que pertence. A PORBASE torna-se assim um rico repositório de referências a obras existentes nas bibliotecas portuguesas, mas a continuação do seu desenvolvimento nestes moldes parece-nos requerer alguma reflexão. Se é verdade que os avanços tecnológicos permitem armazenar cada vez maiores volumes de informação, é igualmente verdade que o desenvolvimento de redes poderá dispensar, em certa medida, repositórios não especializados de informação, que são característica de catálogos colectivos como a PORBASE. Sem pôr em causa a importância desta como recurso informativo, devemos analisar os vários aspectos da sua gestão para melhor perspectivarmos os caminhos a seguir. Com efeito, a continuação da integração de registos tal como se tem realizado até agora coloca, no caso concreto da PORBASE e do sistema que a suporta, alguns tipos de problemas. Em primeiro lugar, ela obriga ao desenvolvimento e constante actualização de aplicações informáticas que permitam avaliar, da forma mais automática possível, a qualidade dos registos que são enviados pelas bibliotecas cooperantes em termos de cumprimento das regras de descrição definidas e que possibilitem um eficaz controlo de

duplicados que evite a sua integração na base. Por muito automatizados que estes processos sejam, eles não evitam o envolvimento humano que é necessário se quisermos garantir elevados níveis de qualidade. Em segundo lugar, a integração de registos cujas entradas de autoridade diferem das formas estabelecidas na base de autoridades a que a base bibliográfica está ligada afectam a consistência desta, ao gerarem frequentemente novas autoridades que correspondem a formas não autorizadas, criando situações que só a intervenção humana pode resolver. Em terceiro lugar temos o problema que se cria quando uma biblioteca cooperante decide alterar todas, ou um substancial número das suas cotas – neste caso, encontrou-se a solução de acrescentar um novo campo na base de existências, com a designação de “cota antiga”, já que se tornava inviável alterar na base todas as cotas modificadas; contudo, embora tal não constitua um obstáculo grave à recuperação da informação, a PORBASE deixará, neste caso, de espelhar com exactidão as existências dessa biblioteca, sendo necessário que o utilizador procure, na respectiva base de dados ou *in loco*, os elementos referentes à sua localização física. Um quarto problema surge quando há necessidade de apagar registos correspondentes a obras que, por qualquer motivo, as bibliotecas deixaram de possuir no seu acervo: no caso de o registo existente na PORBASE ser apenas dessa biblioteca e não se encontrar ligado a nenhum registo de autoridade que seja comum a outros registos bibliográficos, a operação é relativamente simples; no caso de o registo bibliográfico ser comum a várias bibliotecas tem de se proceder à alteração a nível do ficheiro de existências; e no caso em que o registo só existe numa biblioteca mas tem entradas de autoridade (autor, assunto, classificação) comuns a outros registos, a eliminação da existência tem de ser precedida do corte das ligações com a respectiva autoridade. Estas operações não são complexas, mas à medida que a base cresce aumenta também o número de situações para as quais é necessária a intervenção humana, o que significa uma sobrecarga a nível da gestão da base e pode inviabilizar a sua manutenção em termos de actualidade permanente. Pensamos que o desenvolvimento da PORBASE tem sido por vezes confundido com o seu crescimento físico, isto é, com o aumento de número de referências (mesmo que sejam apenas a nível do número de exemplares recenseados) na base central. Sem condenarmos, à partida, esta

visão, parece-nos contudo que ela é bastante redutora e pode até conduzir a situações de difícil retorno, pelas razões que acima apontámos; para uma visão realista do que possa vir a ser o futuro da PORBASE, é necessário que se reflecta sobre os seus propósitos enquanto Catálogo Colectivo em Linha das Bibliotecas Portuguesas. Como foi salientado nas 8^{as} Jornadas PORBASE, em 2000, uma das estratégias a seguir seria “uma redefinição mais precisa do conteúdo da PORBASE, optando-se pela constituição tão exhaustiva quanto possível de uma grande base de dados de bibliografia portuguesa ou referente a Portugal, corrente ou retrospectiva, produzida em circuito comercial ou designada “literatura cinzenta”, em documento primário ou secundário e independentemente do suporte em que esteja disponível” (Campos, 2000). Partindo do pressuposto de que a bibliografia corrente continuaria a ser garantida pela Biblioteca Nacional, através do Depósito Legal, apontavam-se os fundos antigos, os artigos em publicações periódicas, a “literatura cinzenta” e os recursos electrónicos como áreas para as quais se deveriam definir estratégias de partilha de responsabilidades catalográficas (Campos, 2000). Nessa linha, foram apresentados no final de 2003, nas 9^{as} Jornadas PORBASE, dois projectos a desenvolver em cooperação: o projecto Patrimonia – Livro Antigo e o Projecto de Analíticos de Publicações Periódicas Científicas Portuguesas¹³⁶. Em relação à bibliografia estrangeira considerada como um importante contributo para a PORBASE, o caminho seria estimular o acesso a bases de dados estrangeiras que as tivessem referenciadas, com eventual aproveitamento de registos (Campos, 2000).

Se nos referimos às estratégias de desenvolvimento da PORBASE é por considerarmos que elas se prendem inevitavelmente com as questões da recuperação por assunto de que temos vindo a tratar. Com efeito, consideramos que os avanços na área tecnológica devem ser aproveitados com um máximo de eficiência, procurando que se tornem um meio para otimizar o acesso à informação e para rentabilizar o trabalho e o esforço intelectual empregues na produção, na análise e no processamento dessa

¹³⁶ JORNADAS PORBASE, 9, Lisboa, 2003 - *Novos contextos para a cooperação: comunicações*. [Em linha]. [Consult. 23 Mar. 2005]. Disponível em: <URL: <http://purl.pt/331/programa.html>>.

informação. O sistema que suporta a PORBASE implementa o protocolo Z39.50 e esse facto, embora não seja suficiente para que a interoperabilidade exista de forma completa e efectiva, abre-lhe caminho e é uma via possível para aceder à informação existente nas bases de dados de outras bibliotecas portuguesas (e também estrangeiras) que implementem o protocolo, sem necessidade de integrar todos os seus registos na PORBASE, permitindo a selecção e eventual extracção dos que forem considerados relevantes para alimentar a base central. Esta última situação, a concretizar-se, implicaria uma maior atenção aos conteúdos da PORBASE, à sua organização e tratamento temático, e modificaria, em parte, o modelo de cooperação que tem sido constante ao longo da existência da Base Nacional de Dados Bibliográficos. Conforme atrás referimos, é erróneo pensar que a interoperabilidade se resolve apenas no nível da tecnologia. Como referiu William Moen, “tal como o termo *standard*, a interoperabilidade tem vários matizes e níveis de significado” e “o facto de os sistemas poderem interoperar não significa que as organizações queiram os seus sistemas e a informação neles existente acessíveis a alguém” (2000, p. 131). A implementação de uma norma como o Z39.50, ou de qualquer outra que providencie uma infraestrutura técnica que permita a interoperabilidade, tem de ser acompanhada pelo entendimento do que ela implica a nível do acesso à informação, da partilha efectiva de recursos e de uma recuperação fiável e consistente.

Se nos centrarmos na missão e no objecto da PORBASE (ser o catálogo colectivo das bibliotecas portuguesas e especializar-se na bibliografia produzida em Portugal, por autores portugueses e sobre Portugal), poderemos definir níveis de intervenção diferentes, de acordo com os objectivos a atingir. A ligação a catálogos de bibliotecas, nacionais e estrangeiras, através de protocolos como o Z39.50 ou do SRW, poderá constituir uma forma de promover a real partilha de responsabilidades catalográficas a nível da PORBASE, já que a gestão das bases de dados será sempre da responsabilidade dos seus detentores, não havendo necessidade de qualquer intervenção directa a esse nível por parte do Órgão Coordenador. Nesta óptica, podemos encarar a PORBASE como um catálogo virtual de facto, que funciona simultaneamente como um portal através do qual se acede quer a informação

fisicamente integrada numa mesma base quer a um mundo de recursos distribuídos. Mas, para que tal possa acontecer, há todo um trabalho a realizar no sentido de implementar a interoperabilidade e expandi-la do nível tecnológico e descritivo (compatibilidade de formatos) a outros níveis de cooperação, no qual se encontram os aspectos inerentes aos conteúdos temáticos das bases de dados, à sua organização, representação e acesso. Recordando o que dissemos no capítulo 3 sobre as actuais possibilidades de aceder por assunto ao conteúdo informativo da PORBASE, convém reter os seguintes aspectos essenciais:

1. A CDU é usada como linguagem comum de representação dos assuntos, embora muitas bibliotecas a utilizem apenas para a finalidade da cooperação; existem vários níveis de profundidade de utilização da classificação, ficando-se muitas bibliotecas pelos níveis mais elevados da hierarquia, sem usar auxiliares, enquanto outras adoptam um nível de descrição mais profundo e utilizam, sempre que necessário, os vários auxiliares.
2. A linguagem de indexação criada a partir do SIPORbase é aplicada pela Biblioteca Nacional (que é também o maior cooperante da PORBASE) na descrição dos seus recursos, ainda que não de uma forma consistente (há áreas temáticas que não são cobertas e por vezes, devido a outras prioridades, há suspensão temporária da indexação, usando-se apenas a CDU).
3. Existem na base registos indexados com terminologias já aprovadas no âmbito do Projecto CLIP, mas estas estão ainda longe de cobrir todas as áreas temáticas.

Após a análise que fizemos dos três projectos (RENARDUS, MACS e HILT), ressalta a convicção de que os caminhos para a interoperabilidade efectiva a nível dos assuntos andam ainda a ser desbravados, sendo os projectos oportunidades de experimentação e aprendizagem que podem vir a dar origem a verdadeiros serviços (como o RENARDUS que, pelo menos a nível de estrutura de gestão, se encontra já definido). As propostas que aqui apresentamos para o desenvolvimento da PORBASE no que respeita às possibilidades de acesso aos seus conteúdos temáticos inspiram-se em parte nessa análise mas também noutros trabalhos que têm vindo a ser realizados e divulgados nos últimos anos e apontam para estratégias diferentes de acordo

com as várias facetas que a PORBASE poderá apresentar enquanto catálogo colectivo e portal dos recursos existentes em bibliotecas e outros serviços de informação portugueses (ou estrangeiros, sobre Portugal). Seria demasiado ambicioso, e sairia mesmo do âmbito do presente trabalho, pretendermos apresentar soluções, pois temos consciência dos constrangimentos existentes a vários níveis, nos quais se incluem as limitações de ordem financeira, a escassez de recursos humanos especializados, pressões de carácter institucional e factores decorrentes de uma cultura organizacional que é lenta na mudança, que dificultam a implementação de determinadas propostas e obrigam à sua adaptação em função da realidade e do ambiente que envolve as organizações. Mas sejam quais forem os caminhos a seguir, é imprescindível que se conheça o universo dos serviços com quem se trabalha e com quem se pretende vir a trabalhar no futuro. Nessa óptica, parece-nos que, a nível metodológico, o primeiro passo a dar seria a realização de um questionário com o objectivo específico de (re)conhecer as políticas de indexação, as linguagens utilizadas e os modelos da organização da informação nas entidades-alvo, assim como auscultar a sensibilidade destas e dos seus utilizadores em relação às questões do tratamento e recuperação dos assuntos, à semelhança do que foi feito pelo projecto HILT. Partindo do pressuposto que a comunidade de cooperantes da PORBASE deverá ser alargada a outras entidades que não só as bibliotecas (o que levou inclusivamente à criação de um novo tipo de cooperante – o designado “cooperante externo” – aprovado na Assembleia de Cooperantes de Dezembro de 2003¹³⁷), abrindo caminho à diversificação dos recursos disponibilizados a partir de pontos de acesso comuns, com destaque para o acesso por assunto, este questionário teria como universo não só os cooperantes efectivos da PORBASE (bibliotecas) mas também outras entidades que utilizam e desenvolvem sistemas de organização do conhecimento, como museus, arquivos e instituições públicas (como, por exemplo, o IPPAR¹³⁸ e o IPM¹³⁹, na área do património cultural). Este levantamento permitiria enquadrar o problema, perceber quais as dificuldades e obstáculos a ultrapassar,

¹³⁷ Ver <URL: <http://www.porbase.org>>.

¹³⁸ Instituto Português do Património Arquitectónico.

¹³⁹ Instituto Português de Museus.

reconhecer os pontos fortes e fracos do tratamento temático, definir objectivos e delinear a(s) estratégia(s) a seguir. Sem o estudo dos resultados de um inquérito deste tipo dificilmente se podem desenvolver soluções realistas, que envolvam de facto todos os participantes. Como aconteceu na primeira fase do projecto HILT, foi o levantamento da situação em cada comunidade-alvo do projecto que forneceu as bases para as propostas dos futuros desenvolvimentos do projecto.

Entendemos que as instituições, mesmo que não estejam vocacionadas especificamente para a investigação, podem e devem funcionar, até certo ponto (e nunca pondo em causa a sua verdadeira missão e os seus objectivos), como laboratórios de experimentação de novas técnicas e novos processos. Constituem-se assim como organizações “que aprendem”, isto é, como organizações que desenvolvem e renovam o conhecimento. Para isso é necessário que encontrem os meios mais eficientes para realizar essas experiências, recorrendo ao próprio conhecimento organizacional, à experiência e saber daqueles que trabalham nas organizações, desenvolvendo parcerias e definindo metodologias que possam testar a aplicabilidade e a validade de determinadas ideias e hipóteses, de uma forma coordenada. Ainda que o objectivo seja, em geral, a alteração ou melhoria de procedimentos, a renovação ou a criação de novos serviços, é conveniente testar a viabilidade de metodologias, técnicas e processos num universo delimitado, antes de se avançar para um mais global. Daí o interesse dos projectos-piloto, como vimos em relação aos projectos RENARDUS e MACS. Em relação à PORBASE, e porque o ambiente informativo em que nos inserimos deve ser bem conhecido antes de se tomarem decisões sobre um caminho a seguir, é necessário desenvolver, com parceiros concretos e tendo em consideração os recursos existentes, a nível material e humano, projectos-piloto que permitam testar métodos e técnicas e adequá-los à realidade mais ampla que se pretende abranger, de modo a permitir uma planificação o mais realista possível de produtos a desenvolver, serviços a implementar e áreas de cooperação a explorar. Para um melhor enquadramento das várias facetas que o acesso por assunto assume num catálogo colectivo como a PORBASE, organizámo-las

em três grandes grupos: desenvolvimento das linguagens documentais, acesso multilíngue e organização geral do conhecimento na PORBASE.

A nível do desenvolvimento das linguagens documentais, da sua harmonização, compatibilização e adequação ao novo ambiente de rede, poderiam seguir-se os seguintes caminhos:

- Eleger, com base nos resultados do questionário atrás referido, um grupo constituído pela Biblioteca Nacional, por bibliotecas cooperantes e por outras entidades seleccionadas para o efeito, preferencialmente de comunidades diferentes (museus, por exemplo), que desenvolvesse um projecto-piloto que testasse, comparativamente, a eficiência e a eficácia do mapeamento entre linguagens de indexação diferentes (de acordo com o modelo experimentado no MACS para linguagens em diferentes línguas) e de um mapeamento que usasse a CDU como estrutura central, à semelhança do modelo seguido pelo projecto RENARDUS, rentabilizando o facto de o ficheiro de autoridade de assuntos da BN contemplar as equivalências entre os termos de indexação e a notação CDU. Tal como no projecto MACS, seleccionar-se-ia uma (ou mais) área(s) temática(s) sobre a qual incidiria o projecto-piloto. Este trabalho permitiria perceber as reais dificuldades e implicações de um mapeamento entre linguagens documentais diferentes, assim como a sua viabilidade. Mais tarde, caso se verificasse a exequibilidade dos métodos, técnicas e processos seguidos, o projecto poderia alargar-se, ainda que provavelmente de forma faseada, a todo o universo da PORBASE.

- Seleccionar um (ou vários) dos grupos de trabalho do Projecto CLIP para desenvolver ontologias, baseadas nas terminologias já por eles aprovadas, que pudessem vir a ser utilizadas em sistemas de organização do conhecimento com possibilidade de indexação e recuperação automáticas. Numa fase piloto, poderiam seguir-se duas vias paralelas: rentabilizar o facto de o grupo da área temática "Agricultura e Pecuária" coincidir com o grupo de trabalho do tesouro Agrovoc e de o seu coordenador ser o representante nacional do sistema AGRIS, acompanhando e, eventualmente, assimilando as experiências que estão a desenvolver-se no âmbito da FAO e que consistem na transformação e integração em ontologias das várias linguagens existentes para aquelas áreas temáticas e outras afins (como a Pesca); outra via seria seleccionar uma

linguagem criada no âmbito dos grupos de Belas Artes ou do Ambiente (casos em que as linguagens existentes foram desenvolvidas pelos próprios grupos) e utilizá-las para o mesmo fim.

Qualquer destes projectos teria de se desenvolver com a participação empenhada das entidades envolvidas e teria de ser forçosamente interdisciplinar, reunindo bibliotecários e outros profissionais da informação com sólidos conhecimentos em linguagens documentais, informáticos e especialistas das áreas temáticas abrangidas.

Ao nível do acesso multilíngue, seria interessante integrar, numa primeira fase a título experimental e numa área temática restrita, o projecto MACS, com a linguagem criada na Biblioteca Nacional a partir do SIPORbase e validada pelo CLIP. Seria um trabalho que potenciaría as possibilidades de pesquisa dos recursos existentes na PORBASE, alargando-a a nível internacional – o que, estrategicamente, poderia constituir um novo motivo de interesse e mais uma razão para as bibliotecas portuguesas aderirem à PORBASE e utilizarem vocabulários normalizados na indexação.

A nível da organização de conhecimentos na PORBASE, a questão coloca-se a vários níveis, que se prendem com as características actuais da base de dados e do sistema que a suporta. Tal como acontece com outros catálogos do mesmo tipo, que reúnem dados produzidos por várias bibliotecas e em diferentes épocas, temos registos correspondentes a documentos classificados de acordo com diferentes versões da CDU, apenas classificados, classificados e indexados ou, pura e simplesmente, sem qualquer ponto de acesso por assunto. Se algumas situações podem ser resolvidas automaticamente (no caso das notações da CDU que foram objecto de alteração, por exemplo, mas apenas quando as notações estão atribuídas na sua forma mais simples, sem qualquer tipo de auxiliar, o que corresponde a uma percentagem mínima de ocorrências), outras teriam de ser objecto de trabalhos de conversão retrospectiva, morosos e envolvendo uma forte componente humana, mesmo que os registos não fossem alterados e se optasse por estabelecer um quadro de equivalência entre as notações antigas e as mais recentes, com base no qual seriam lançadas as pesquisas na base. Pese embora o facto de, como

vimos, muitas bibliotecas utilizarem a CDU apenas para fins de cooperação na PORBASE e de haver alguma pressão no sentido de se aceitarem registos sem aquele sistema de representação, tem sido internacionalmente reconhecida a vantagem das linguagens estruturadas e com elevado nível de formalização (sejam elas notacionais ou terminológicas) na organização do conhecimento. Assim, a manutenção da CDU como linguagem comum na Base Nacional de Dados Bibliográficos parece-nos ser de manter, pelas razões que foram apontadas aquando da criação da PORBASE e que se mantêm válidas ainda hoje – é uma linguagem estruturada, normalizada, internacional e que cobre todas as áreas temáticas, permitindo, ainda que com limitações, o acesso aos assuntos referenciados na base. Possui, além disso, potencialidades ainda não completamente exploradas, mas que o poderão ser quando forem implementadas algumas melhorias a nível do MRF e do formato Unimarc/Classificação.

Um outro recurso que deveria ser rentabilizado e aproveitado em todas as suas potencialidades é o ficheiro de autoridades de assunto, construído à medida que são indexados os documentos na Biblioteca Nacional e a que se encontram ligados os respectivos registos bibliográficos na PORBASE. O formato UNIMARC/Autoridades em que assenta permite a utilização de códigos detalhados para os elementos dos vocabulários controlados e pode ser usado para estabelecer um sistema de equivalência entre linguagens documentais diferentes. Na Biblioteca Nacional essa equivalência é feita entre o cabeçalho de assunto e a CDU, mas os registos podem, se necessário, ser mais desenvolvidos, à semelhança do que se fez para o mapeamento entre os LCSH, a CDD e outras linguagens, em que os vocabulários a mapear foram primeiro convertidos no formato de autoridade MARC 21 (Vizine-Goetz et al., 2004). A disponibilidade pública do ficheiro de autoridades, a par da implementação do formato Unimarc/Autoridades por outras bibliotecas ou entidades afins, poderá constituir mais um aspecto a desenvolver na área da cooperação, promovendo uma real gestão e partilha de recursos.

Num dos projectos a desenvolver no âmbito da PORBASE – o dos analíticos de publicações periódicas científicas portuguesas¹⁴⁰ – que prevê a ligação aos sumários de cada número e, quando possível, aos resumos dos artigos, poderão explorar-se as potencialidades da pesquisa em texto livre e testar sistemas de indexação e recuperação automáticas. Poderão, assim, aplicar-se critérios diferenciados no tratamento dos assuntos, contemplando uma descrição de conteúdos temáticos ao nível de colecção (considerando as colecções, para este efeito, agrupamentos de assuntos com alguma homogeneidade), para o que nos parece apropriado utilizar a estrutura da CDU e, se possível, uma descrição de conteúdos a nível do item, mais detalhado, de acordo com sistemas e terminologias de indexação normalizados e em consonância com o que for decidido pelos parceiros do projecto.

No catálogo físico comum, o esforço deveria ir no sentido da indexação sistemática e consistente de núcleos de informação seleccionados de bibliografia portuguesa, de autores portugueses e sobre Portugal, para além da utilização da CDU como linguagem geral de descrição e recuperação a aplicar a todos os registos. Aos cabeçalhos construídos a partir dos princípios e regras do SIPORbase, resultantes sobretudo da indexação feita pela Biblioteca Nacional, deveriam juntar-se, por parte das bibliotecas cooperantes, as terminologias aprovadas no âmbito do Projecto CLIP, o que pressupõe a reactivação de alguns dos seus grupos de trabalho que, pelas mais diversas razões, têm a actividade suspensa. Um dos problemas desta iniciativa é o seu carácter mais ou menos voluntário, pois embora exista um compromisso de colaboração por parte das instituições a que os elementos dos grupos temáticos pertencem, o facto de estes terem de acumular a actividade no CLIP com a sua actividade profissional corrente dificulta por vezes o trabalho conjunto e leva a que os resultados sejam mais lentos do que seria desejável. Por outro lado, as constantes mudanças a nível institucional têm afectado, em

¹⁴⁰ Veja-se, sobre este projecto, LOPES, Maria Margarida - Analíticos de publicações periódicas científicas portuguesas: um projecto de cooperação no âmbito da PORBASE. JORNADAS PORBASE, 9, Lisboa, 2003 - *Novos contextos para a cooperação*. Lisboa: BN, 2003.

certos casos, a estabilidade dos grupos de trabalho é dificultada a sua actividade.

A nível global, perspectivando a PORBASE como catálogo colectivo distribuído, pensamos que uma das questões mais importantes será a uniformização dos pontos de acesso, sejam eles a nível de autores, sejam a nível de assuntos. Neste último caso, e considerando a existência de bases de dados organizadas de acordo com tipologias documentais (Livro Antigo, analíticos, a própria BND) independentemente de os registos que as compõem estarem ou não referenciados na base central, impõe-se um acesso único que permita a recuperação simultânea dos itens, seja qual for a base em que se encontram. Um caminho seria estudar a hipótese de organizar os recursos em colecções temáticas, que poderiam funcionar como “porta de acesso” a um conjunto consistente de recursos em qualquer suporte e independentemente da sua localização física - no catálogo colectivo central, na Patrimonia, na base de analíticos ou noutra. Esta descrição ao nível de colecção poderia basear-se na CDU, com equivalência desejável numa linguagem terminológica controlada, aos níveis mais elevados da sua hierarquia (a definir, muito provavelmente com diferentes opções a nível das subclasses para as diferentes classes da CDU, já que se teria de ter em conta as particularidades da organização da tabela relativamente a cada uma delas). Num estágio mais avançado, esta linguagem terminológica controlada, assim como a CDU, seria a mediadora entre o conteúdo temático global de cada agregado de informação e a linguagem natural do utilizador, o que pressupõe a existência de uma estrutura “oculta” que traduza esta última na linguagem compreendida pela máquina - seja ela uma ontologia ou os elementos de um ficheiro de autoridade.

Sem a pretensão de ter esgotado o assunto e considerando existirem outros caminhos possíveis para o desenvolvimento da Base Nacional de Dados Bibliográficos, que passam, por exemplo, pelo tipo de colaboração que se pode estabelecer com outras iniciativas que têm vindo a desenvolver-se recentemente, como a Biblioteca do Conhecimento (bON), ou por uma ligação a redes de bibliotecas já existentes ou em embrião, que poderão implicar outras estratégias para um efectivo acesso por assunto, pensamos ter,

contudo, apontado alguns dos caminhos possíveis para tornar mais consistente, fiável e lato o acesso à informação na PORBASE.

7. Conclusão

Muitas vezes sentimos, quando chegamos ao final de um trabalho, que estamos então em condições de o iniciar. De facto, à medida que vamos desenvolvendo uma reflexão sobre um determinado tema vamos construindo novas hipóteses de trabalho e são inúmeras as questões levantadas que ficam a aguardar uma análise futura. Fica ainda a sensação de que muito ficou por dizer. Daí que, ao chegarmos a este ponto, seja difícil tirar conclusões e nos pareça mais correcto sintetizar alguns aspectos analisados e deixar algumas reflexões em aberto.

Sem dúvida que os resultados dos vários projectos que se têm vindo a desenvolver, uma vez analisados, poderão apontar vias possíveis para solucionar pelo menos alguns dos problemas existentes na área do acesso por assunto à informação em rede. Tornar-se-á necessário, contudo, integrar esses resultados em propostas exequíveis ao nível de serviços sustentados e não apenas de projectos, tentando inseri-las nas estratégias das organizações que se destinam a ampliar e melhorar o acesso à informação.

Uma questão que nos parece bastante consensual é a vantagem de usar, sempre que possível, como ponto de partida ou como estrutura fundamental, formas de organização e representação do conhecimento já estabelecidas, como as classificações, os tesauros, as listas de cabeçalhos de assunto e outros instrumentos terminológicos usados para indexar documentos, potenciando assim todo o trabalho e investimento que, ao longo dos tempos, tem sido desenvolvido para permitir um melhor acesso à informação. O exemplo dos portais temáticos ("*subject gateways*") mostra-nos que a organização do conhecimento e a sua recuperação através de esquemas de classificação já existentes e, de certa forma, familiares aos utilizadores, facilita uma abordagem global ao acervo de informação que disponibilizam. A adaptabilidade destas linguagens à classificação automática de documentos

disponíveis em texto integral foi experimentada num estudo piloto para a base de dados *WorldCat*, da OCLC, a partir de um subconjunto da LCC¹⁴¹ – a experiência demonstrou ser ainda necessária alguma mediação humana para atingir um nível de qualidade mais elevado, pela exigência de maior especificidade. A metodologia seguida – simplificação da tabela da LCC, mapeamento entre as classes remanescentes da simplificação e duas fontes de terminologia (LCSA – *Library of Congress Subject Authority File*¹⁴² - e a base de dados bibliográficos *WorldCat*¹⁴³) – demonstrou a necessidade de algumas adaptações e desenvolvimentos, embora os resultados fossem considerados satisfatórios para o acesso a recursos que não tinham outra possibilidade de recuperação por assunto.

A indexação, “linguística, conceptual, estrutural, [...] continua ainda na frente de cena no muito actual projecto da *Web* semântica, graças às novas formas de representação que são as ontologias” (Amar, 2004, p. 61). A continuação do desenvolvimento de ontologias permitirá a criação de instrumentos consistentes, adequados a uma utilização por computadores, que expandam, afinem e estructurem, de modo mais consolidado e inequívoco, os tradicionais tesouros, introduzindo igualmente características existentes em alguns sistemas de classificação. São instrumentos normalizadores e estruturantes das linguagens documentais e podem tornar-se importantes meios para indexar o conhecimento e permitir uma maior automatização do processo.

Os mapeamentos entre linguagens já existentes, sejam eles de equivalência entre notações e termos de indexação, ou entre notações de sistemas de classificação diferentes (como se viu no caso do projecto RENARDUS), ou de equivalência entre diferentes termos de indexação terminológica (como vimos, em relação a diferentes línguas, no projecto MACS), parecem apresentar-se como um caminho igualmente promissor. A nível metodológico, esses mapeamentos podem desenvolver-se de duas formas distintas: optando pela igualdade entre as linguagens, estabelecendo equivalências entre elas e

¹⁴¹ Veja-se Godby e Stuler (2001).

¹⁴² <URL: <http://authorities.loc.gov/>>

¹⁴³ <URL: <http://www.oclc.org/worldcat/default.htm>>

respeitando a sua autonomia, como no projecto MACS, ou escolhendo uma linguagem de comutação (ou “estrutura central”) à qual as outras linguagens e esquemas são mapeados (como no projecto RENARDUS e na proposta do HILT, com a CDD).

Da análise que realizámos sobre os catálogos colectivos nas suas diversas facetas (estrutura da informação, sistemas que os suportam, pontos de acesso), sobre os projectos RENARDUS, MACS e HILT e sobre a PORBASE, ressalta um aspecto quanto a nós fundamental, que é a necessidade de integrar os resultados de todas as experiências e iniciativas em serviços onde a fluidez dos processos seja uma evidência. É importante que as soluções encontradas se destinem a uma utilização colectiva, independentemente das comunidades que as utilizem e dos sistemas que estas possuam. A procura dessas soluções deve ter sempre em consideração a interoperabilidade, quer a nível tecnológico, quer a nível dos conteúdos e das próprias relações inter-organizacionais, através do estabelecimento de acordos, princípios, regras, procedimentos e normas comuns. Esta atitude obriga, por vezes, a adaptar os mecanismos de cooperação já existentes ao mesmo tempo que conduz à criação de novas áreas de cooperação e a novos parceiros.

A nível do acesso por assunto nos novos ambientes de rede, não é demais salientar a enorme vantagem em aproveitar e potenciar todo o esforço que tem sido desenvolvido pelas bibliotecas e entidades afins no tratamento temático dos conteúdos dos seus acervos e todo o trabalho intelectual investido na construção de instrumentos de representação de assuntos. Uma vez que cada unidade de informação possui os seus sistemas de organização do conhecimento, que reflectem a sua realidade individual ainda que se apoiem em sistemas formais, normalizados e genericamente aceites, quando pretendem ligar-se a outras unidades há sempre diferenças que têm de ser resolvidas ou ultrapassadas para que o acesso à informação que disponibilizam seja efectivo. Se a transferência de dados entre diferentes sistemas tem sido considerada um ponto fulcral, existem outros aspectos igualmente problemáticos, mesmo em serviços que utilizam a mesma linguagem de representação de assuntos. As diferenças detectadas a nível da

profundidade da análise e representação de assuntos, por exemplo, é um problema difícil de resolver, já que ele tem a ver com políticas de indexação e decisões relativas aos vários tipos de utilizadores da informação, o que nos conduz à questão, sempre em aberto, de saber se a indexação deve basear-se mais nos conteúdos da informação ou nos utilizadores a que se destina. Na realidade, parece-nos que o equilíbrio entre ambas as posições é a resposta mais adequada. O que indexamos é o conteúdo dos documentos e essa indexação deve ser a mais específica possível, ainda que um utilizador menos especializado possa não aceder facilmente à informação assim tratada. Contudo, as cadeias semânticas em que os termos se inserem, por um lado, ou a utilização simultânea de uma linguagem mais lata que permita enquadrar o assunto, por outro (à semelhança do que acontece na Biblioteca Nacional, quando se utiliza uma terminologia especializada e constroem cabeçalhos específicos segundo as regras do SIPORbase e, paralelamente, se categoriza essa informação numa classe CDU), permitem a utilizadores com diferentes graus de exigência ou de conhecimentos aceder à mesma informação. As próprias linguagens documentais são diferentes quanto ao nível de profundidade com que tratam os assuntos, mesmo quando pertencem à mesma tipologia (classificações, tesouros...), como verificámos. Os mapeamentos que o projecto RENARDUS e MACS desenvolveram tiveram em consideração essas diferenças, também existentes entre as várias linguagens que analisaram, de modo a aumentar os níveis de revocação na recuperação da informação.

No caso concreto da PORBASE, que tem utilizado a CDU como linguagem comum de representação e recuperação de assuntos, referimos também que as bibliotecas utilizam diferentes níveis da hierarquia e desenvolvem de modo desigual as notações para representar os mesmos assuntos, o que poderá apontar a necessidade de definir alguns critérios comuns mínimos de utilização desta classificação, tentando normalizar ao máximo a construção das notações.

Interoperabilidade, normalização, cooperação, investigação, integração – parecem-nos palavras-chave essenciais no actual estado do acesso por

assuntos em ambientes de rede. A interoperabilidade exige normalização e a cooperação é essencial na elaboração de normas, protocolos e outros *standards* que possam ser aceites por vastas comunidades. Os resultados da investigação e dos projectos desenvolvidos nesta área devem ir sendo testados a um nível mais alargado, abrangendo mais organizações e serviços, a fim de se verificar a sua eficácia e a possibilidade de integração em diferentes sistemas. Para que se amplie e diversifique a informação disponibilizada e se atinjam melhores níveis de qualidade. Para que um dia, de facto, possamos dizer: "*The librarian is in the strange but fortunate position of being able to say: We have the answers - what are the questions?*"¹⁴⁴

¹⁴⁴ FOSKETT, A.C. – *The subject approach to information*, 4ª ed. London: Clive Bingley, 1982, p. 84.

Bibliografia¹⁴⁵

AMAR, Muriel (2004) – L'indexation aujourd'hui. *Les dossiers de l'ingénierie éducative*. N.º 49 (Déc. 2004), p. 61-65.

AMERICAN LIBRARY ASSOCIATION (1997) – *Final report to the ALCTS/CCS subject analysis committee - subcommittee on subject relationship / reference structures* [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.ala.org/alcts/organization/ccs/sac/rpt97rev.html>>

ANSI/NISO Z39.59. 2003 – *Information retrieval (Z39.50): application service definition and protocol specification*. Bethesda, Maryland: NISO. IX+267 p. ISSN 1041-5653 [Em linha]. [Consult. 30 Nov. 2004]. Disponível na WWW: <URL: <http://www.niso.org/standards/resources/Z39-50-2003.pdf>>

ARDÖ, Anders [et al.] ([1994]) – *Improving resource discovery and retrieval on the internet : the Nordic WAIS / World Wide Web Project : summary report*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.lub.lu.se/W4/summary.html>>.

ARDÖ, Anders [et al.] (2001) – *Specification of Renardus Pilot broker system : Renardus deliverable D2.3 version 1.3*. 2001 [Em linha]. [Consult. 14 Out. 2002]. Disponível na WWW: <URL: <http://www.renardus.org/>>

BACA, Martha – Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information. *Cataloging & Classification Quarterly*. Vol. 36, N.º 3-4 (2003). p. 47-55.

BASILI, Carla (1998) – *La biblioteca in rete : strategie e servizi nella società dell'informazione*. Milano: Editrice Bibliografica, 1998. ISBN 88-7075-503-7.

¹⁴⁵ A bibliografia foi construída de acordo com a NP 405-1. 1994, Informação e Documentação – *Referências bibliográficas: documentos impressos*. Lisboa: IPQ. 49 p. e com a NP 405-4. 2001, Informação e Documentação – *Referências bibliográficas: parte 4: documentos electrónicos*. Lisboa: IPQ. 28 p.

BERNERS-LEE, Tim (2001) – *Business model for the semantic Web*. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://www.w3c.org/DesignIssues/Business>>

BERNERS-LEE, Tim [et al.] (2004) – *Frequently asked questions about RDF*. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://www.w3.org/RDF/FAQ>>

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora (2001) - *The Semantic Web, Scientific American*. May 2001. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>

BIBLIOTECA NACIONAL. Grupo de Implementação do Sistema Horizon (2000) – PORBASE : Novas perspectivas de desenvolvimento. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000].

BINDING, Ceri; TUDHOPE, Douglas (2004) – KOS at your service : programmatic access to knowledge organisation systems. *Journal of digital information*. Vol. 4, N.º 4, Artigo n.º 255 (2004.02.05). [Em linha]. [Consult. 16 Mar. 2005]. Disponível na WWW: <URL: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/>>

BORBINHA, José (2003) – L'authority control nel mondo dei metadati. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

BOURDON, Françoise (2003) – Elaborare i modelli dei dati d'autorità per le biblioteche, gli archivi e i musei: tentativo in corso all'AFNOR. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control :definition*

and international experiences. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

BUIZZA, Pino (2003)– Controllo bibliografico e authority control dai Principi di Parigi a oggi. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

CABRAL, Maria Luísa (1987) – A informatização da Biblioteca Nacional : gerir o presente, planear o futuro. *Revista da Biblioteca Nacional*. 2ª série, vol. 2, nº1 (1987), p. 161-167.

CAIRNS : *Co-operative Academic Information Retrieval Network for Scotland*. [Em linha]. [Consult. 21 Mar. 2005]. Disponível na WWW: <URL: <http:// Cairns.lib.gla.ac.uk/>>

CAMPOS, Fernanda Maria Guedes de (1995) – 100 bibliotecas na ponta dos dedos : para um conceito de biblioteca virtual. In JORNADAS PORBASE, 6, Lisboa, 1995 – [Comunicações]. Lisboa: Biblioteca Nacional, 1995.

CAMPOS, Fernanda Maria (2000) - Os caminhos da PORBASE. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas do novo milénio*. Lisboa: Biblioteca Nacional, [2000].

CAMPOS, Fernanda Maria (2002) – Breve introdução aos metadados. *Metadados e sua relevância para as bibliotecas*. Lisboa: Biblioteca Nacional, 2002.

CARVALHO, Joaquim de; CORDEIRO, Maria Inês (2002) – XML and bibliographic data : the TVS (Transport, Validation and Services) model. In IFLA COUNCIL AND GENERAL CONFERENCE, 68, Glasgow, 2002 – [Proceedings] [Em linha]. [Consult. 14 Ago. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla68/papers/075-095e.pdf>>

CHAN, Lois Mai; ZENG, Marcia Lei (2002) – Ensuring interoperability among subject vocabularies and knowledge organization schemes : a methodological analysis. In IFLA COUNCIL AND GENERAL CONFERENCE, 68, Glasgow, 2002 – [*Proceedings*] [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla68/papers/008-122e.pdf> >

CHAPMAN, Ann; DAY, Michael; HIOM, Debra (1998) – Metadata : cataloguing practice and Internet subject-based information gateways. *Ariadne*, Issue 18 (Dez. 1998). ISSN 1361-3200. [Em linha]. [Consult. 20 Fev. 2003]. Disponível na WWW: <URL: <http://www.ariadne.ac.uk/issue18/metadata/intro.html>>

CHRYSOS, Thanasis – *Gold: a thesaurus browsing system* [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.csd.ucl.ac.uk/~chrysos/diploma/abstract.html>>

CLARK, Judith (2001) – Subject Portals. *Ariadne*. ISSN 1361-3200. N.º 29 (Set 2001) [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.ariadne.ac.uk/issue29/clark/intro.html>>

CLAVEL-MERRIN, Geneviève (2003a) – MACS (Multilingual Access to Subject) : un authority file virtuelle multilingue. In INTERNATIONAL CONFERENCE AUTHORITY CONTROL, Florence, 2003 - *Definition and international experience*. [Em linha]. [Consult. 19 Mai. 2003]. Disponível na WWW: <URL: <http://www.unifi.it/biblioteche/ac/it/programma.htm>>

CLAVEL-MERRIN, Geneviève (2003b) – National libraries as access points : the role of TEL and MACS. In IFLA GENERAL CONFERENCE AND COUNCIL 69, Berlim, 2003 - *World Library and Information Congress* [Em linha]. [Consult. 11 Jun. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla69/papers/028e-Clavel-Merrin.pdf>>

CLIP – Compatibilização de Linguagens de Indexação em Português (1995) – *Projecto CLIP: fontes para o controlo de terminologia*. Lisboa: Instituto da Biblioteca Nacional e do Livro, 1995. ISBN 972-565-206-1.

COMPUTORIZED Bibliographic Record Actions. [Em linha]. [Consult. 27 Jun. 2003]. Disponível na WWW: <URL: http://www.kb.nl/gabriel/projects/en/partnerships_frame_en.html>

CORDEIRO, Maria Inês (2001) – Tecnologias, bibliotecas e arquitecturas de informação: dos sistemas aos objectos. In CONGRESSO NACIONAL [DE] BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 7, Lisboa, 2001 - *Informação, o desafio do futuro: actas*. Lisboa: APBAD, [2001].

CORDEIRO, Maria Inês (2001a) – From library authority control to network authoritative metadata sources. In IFLA SATTELITE MEETING, Dublin (OH), 2001 - *Subject retrieval in a networked environment : proceedings*. München : K.G. Saur, 2003. p. 131-139.

CORDEIRO, Maria Inês; CARVALHO, Joaquim de (2003) - Fronteiras incertas: da integração dos sistemas à expansão dos serviços. *Cadernos BAD*. N.º 2 (2003), p.36-54.

CORREIA, Ana Maria Ramalho; NETO, Miguel de Castro (2001) – *LIZA – Repositório de literatura científica cinzenta, produzida em Portugal : relatório final*. Lisboa: ISEGI.CEGI, 2001. [Em linha]. [Consult. 21 Nov. 2002]. Disponível na WWW: <URL: http://bnd.bn.pt/ed/liza/enquadramento_teorico.html>

CORSON-RICKERT, Jon (2003) – *Ontologies : What, why and how?*. [Em linha]. [Consult. 16 Mar. 2005]. Disponível na WWW: <URL: <http://metadata-wg.mannlib.cornell.edu/forum/2003-04-18/Ontologies.MWG.20030418.ppt#3o>>

COURBIÈRES, Caroline (2002) – Une approche communicationnelle de l'analyse documentaire. In COLLOQUE INTERNATIONAL, Toulouse, 2002 - *Recherches récentes en sciences de l'information : convergences et dynamiques : actes*. Paris: ADBS, 2002. ISBN 2-84365-059-3. p. 105-125.

CRISTÁN, Ana Lupe (2004) – Avances del programa SACO en Latinoamérica. In IFLA GENERAL CONFERENCE AND COUNCIL, 70, Buenos Aires, 2004 –

[*Proceedings*]. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL:<http://www.ifla.org/IV/ifla70/>>

CRISTÁN, Ana L. (2003) – SACO e i subject gateways. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

CROFTS, Nicolas ; LE BOEUF, Patrick ; ARTUR, Odile (2002) – ELAG Presentation : Ontologies. In LIBRARY SYSTEMS SEMINAR, 26, Roma, 2002 - *Semantic web and libraries*. [Em linha]. [Consult. 19 Set. 2002]. Disponível na WWW: <URL: <http://www.kb.nl/coop/elag/elag2002/www.ifnet.it/elag2002/papers/pap9.html>>

DAY, Michael – *Renardus DDC classification mapping : a summary of work in progress*. Junho 2001. [Em linha]. [Consult. 18 Fev. 2005]. Disponível na WWW: <URL: <http://hilt.cdlr.strath.ac.uk/Dissemination/Workshop%20documents/Renardus%20Report.doc>>.

DEAN, Rebecca J. (2003) - FAST : la messa a punto di intestazioni semplificate per i metadati. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

DELSEY, Tom (1996) – L'évolution des formats MARC. *L'avenir des formats de communication*. Ottawa, 1996. [Em linha]. [Consult. 26 Dez. 2001]. Disponível na WWW: <URL: <http://www.acctbief.org/avenir/evmarcF.htm>>

DOERR, Martin (2001) – Semantic problems of thesaurus mapping. *Journal of digital information*. Vol. 1, n.º 8 (Mar. 2001). [Em linha]. [Consult. 21 Out. 2002]. Disponível na WWW: <URL: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>>

DUNSIRE, Gordon (2002) – Joined up indexes : interoperability issues in Z39.50 networks. In IFLA COUNCIL AND GENERAL CONFERENCE, 68, Glasgow, 2002 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla68/papers/022-144e.pdf> >

DUNSIRE, Gordon (2003) – Use of DDC in HILT 2 and beyond. In DEWEY DECIMAL CLASSIFICATION SEMINAR, Boston Spa, 2003 – [*Papers*] [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW: <URL: <http://cdlr.strath.ac.uk/pubs/dunsireg/useddchilt.pps>>

FACET ANALYTICAL THEORY IN MANAGING KNOWLEDGE STRUCTURE FOR HUMANITIES. [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW: <URL: <http://www.ucl.ac.uk/fatks/prtotype.htm>>

FALCOZ, Franck (2000) - Technical standards : Z39.50 to Web gateway. In [*Renardus*] *Deliverable 2.1 (internal) : technical standards*. [Em linha]. [Consult. 10 Ago. 2000]. Disponível na WWW: <URL:http://www.renardus.org/deliverables/D2_1_Technical_standards.htm>

FISSEHA, Frehiwot; LIANG, Anita; KEIZER, Johannes (2003) – Reengineering AGROVOC to ontologies : step towards better semantic structure. In NKOS WORKSHOP, Houston (Texas), 2003 – [*Papers*]. [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW: <URL: <ftp://ext-ftp.fao.org/Agris/Presentations/NKOSWorshopOntology.zip>>

FONTANA, Antonia Ida (2004) – Subject indexing between international standards and local contexts : the Italian case. In IFLA GENERAL CONFERENCE AND COUNCIL, 70, Buenos Aires, 2004 – [*Proceedings*]. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: URL:<http://www.ifla.org/IV/ifla70/>

FOSKETT, A. C. – *The subject approach to information*. 4th. Ed. London: Clive Bingley, 1982.

FUGMANN, Robert (1995) –The complementary of natural and controlled language in indexing. *Subject indexing : principles and practices in the 90's*. München: K. G. Saur, 1995. p. 215-230.

GALVÃO, Rosa Maria (2000) - Desenvolvimento da base de dados : apresentação das actividades e resultados da cooperação. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000].

GARCÍA CAMARERO, Ernesto; GARCÍA MELERO, Luís Angel (2001) – *La biblioteca digital*. Madrid: Arco/Libros, 2001. ISBN 84-7635-486-X.

GARCIA JIMENEZ, Antonio (1999) – Consecuencias de Internet para el estudio del análisis documental y de los lenguajes documentales : estudio de caso sobre el discurso periodístico. *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: ISKO-España; Universidad, 1999, vol. 3. ISBN 84-930072-0-X. ISSN 1.139-2959.

GARDIN, Jean-Claude (1973) – Document analysis and linguistic theory. *The journal of documentation*. Vol. 29, n.º2 (Jun 1973), p. 137-168.

GODBY, Carol Jean; MILLER, Eric J.; REIGHART, Ray R. – *Automatically generated topic maps of World Wide Web resources*. [Em linha]. [Consult. 21 Mar. 2005]. Disponível na WWW:<URL: <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003493>>

GODBY, Carol Jean; STULER, Jay (2001) – The Library of Congress Classification as a knowledge base for automatic subject categorization. In IFLA PRECONFERENCE - *Subject Retrieval in a Networked Environment*, Dublin (Ohio), 2001 – [Proceedings]. [Em linha]. [Consult. 7 Abr. 2004]. Disponível na WWW:<URL: http://staff.oclc.org/~godby/auto_class/godby-ifla.html>

GÖDERT, Winfried (2001) – Knowledge organization and information retrieval in times of change – concepts for education in Germany. In IFLA COUNCIL

AND GENERAL CONFERENCE, 67, Boston, 2001 - [*Proceedings*]. [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla67/papers/022-142e.pdf>>

GORMAN, Michael (2003) – L'authority file nel contesto del controllo bibliografico in ambiente elettronico. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

GUÉDON, Jean-Claude (2001) – *Internet : le monde en réseaux*. Paris : Gallimard, [2001]. ISBN 2-07-053490-1.

HEDEGAARD, Ruth (2003) – Archives, bibliothèques et musées : les bénéfices de la coopération. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: http://www.ifla.org/IV/ifla69/papers/051f_trans-Hedegaard.pdf>

HEERY, Rachel; CARPENTER, Leona; DAY, Michael (2001) – Renardus project developments and the wider digital library context. *D-Lib magazine*. ISSN 1082-9873. April 2001 [Em linha]. [Consult. 22 Out. 2002]. Disponível na WWW: <URL: <http://www.dlib.org/dlib/april01/heery/04heery.html>>

HILT - HILT stakeholder survey : results. *HILT Phase I : reports*. [Em linha]. [Consult. 18 Fev. 2004]. Disponível na WWW: <URL: <http://hilt.cdjr.strath.ac.uk/Reports/SurveyResults.html>>.

HILT WORKSHOP, Glasgow, 2001 - *Report and conclusions*. [Em linha]. [Consult. 21 Maio 2003]. Disponível na WWW: <URL: <http://hilt.cdjr.strath.ac.uk/Dissemination/WorkshopNew.html>>

HILT WORKSHOP, Glasgow, 2001 - *Breakout sessions: discussion issues*. [Em linha]. [Consult. 21 Maio 2003]. Disponível na WWW: <URL:

<http://hilt.cdlr.strath.ac.uk/Dissemination/Workshop%20documents/BrkoutfullIFINALword.doc>>

HJORLAND, Birger – *Information seeking and subject representation*. Connecticut: Greenwood Press, 1997. ISBN 0-313-29893-9.

HOLZEM, Maryvonne (1999) – *Terminologie et documentation : pour une meilleure circulation des savoirs*. Paris : ADBS, 1999. ISBN 2-84365-032-1.

HOPPENBROUWERS, Jeroen, ed. lit. (2001) – *Architecture of the MACS system*. [Em linha]. [Consult. 14 Fev. 2004]. Disponível na WWW: <URL: <http://laborix.uvt.nl/prj/macs/pub/architecture.pdf>>

HOWARTH, Lynne–C. (2003) – Metadata schemas for subject gateways. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [Proceedings]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla69/papers/053e-Howarth.pdf>>

HUDON, Michèle (2001) – Structuration du savoir et organisation des collections dans les répertoires du Web. *Bulletin des bibliothèques de France*. T. 46, n.º 1 (2001), p. 57-62.

HUXLEY, Lesly (2000) – Follow the fox to Renardus : an academic subject gateway service for Europe. In EUROPEAN CONFERENCE, 4 - *Research and advanced technology for digital libraries : proceedings*. Berlin [etc.]: Springer, 2000. ISBN 3-540-41023-6. p. 395-398.

ICOM. CIDOC CRM SPECIAL INTEREST GOUP (2004) – *Definition of the CIDOC Conceptual Reference Model: version 4.0 (April 2004)*. [Paris]: ICOM, 2004. [Em linha]. [Consult. 9 Fev. 2005]. Disponível na WWW: <URL: http://zeus.ics.forth.gr/official_release_cidoc.html>

IFLA (2002) - *Strategic plan for the UNIMARC Core Activity*. IFLA, 2002. [Em linha]. [Consult. 9 Jun. 2003]. Disponível na WWW:

<URL: <http://www.ifla.org/VI/3/annual/spunimarc.htm>>

IFLA. UNIVERSAL BIBLIOGRAPHIC CONTROL AND INTERNATIONAL MARC CORE PROGRAMME – *UNIMARC : an introduction*. [Em linha]. [Consult. 27 Mar. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/VI/3/p1996-1/unimaec.htm>>

ISO 2788.1986, Documentation - *Guidelines for the establishment and development of monolingual thesauri*. [Genève]: ISO, 1986.

KIRRIEMUIR, John [et al.] (1998) – Cross-searching subject gateways : the query routing and forward knowledge approach. *D-Lib magazine*. ISSN 1082-9873. January 1998 [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.dlib.org/dlib/january98/01kirriemuir.html>>

KOCH, Traugott (2000) – Quality-controlled subject gateways : definitions, typologies, empirical overview. *Online information review : the international journal of digital information research end use*. Vol. 24, nº1, 2000:26. [Em linha]. [Consult. 27 Out. 2004]. Disponível na WWW: <URL: <http://www.mcb.co.uk/oir.htm>>

KOCH, Traugott; NEUROTH, Heike; DAY, Michael (2001) – *Renardus : cross-browsing European subject gateways via a common classification system (DDC)*. July 2001 [Em linha]. [Consult. 9 Jul. 2002]. Disponível na WWW: <URL: <http://www.lub.lu.se/~traugott/drafts/preifla-final.html>>

KOCH, Traugott; NEUROTH, Heike; DAY, Michael (2001) – *Renardus : cross-browsing European subject gateways via a common classification system (DDC)*. In IFLA SATTELITE MEETING, Dublin (Ohio), 2001 - *Subject retrieval in a networked environment : proceedings*. München : K.G. Saur, 2003. p. 25-33.

KOCH, Traugott, et al. (1997) – Specification for resource description methods: part 3 : the role of classification schemes in Internet resource description and discovery. *Project RE 1004 (RE)*. *DESIRE – Development of a European*

Service for Information on Research and Education : deliverable 3.2.3. 19 Fev. 1997. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: http://www.ub.lu.se/desire/radar/reports/D3.2.3/class_v10.html>.

KRAUSE, Jürgen (2003) – Standardization, heterogeneity and the quality of content analysis : a key conflict of digital libraries and its solution. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: http://www.ifla.org/IV/ifla69/papers/085e_trans-krauze.pdf>

LAHARY, Dominique (2002) – Des bibliothèques traditionnelles aux « bibliothèques virtuelles ». *La recherche d'information sur les réseaux*. Paris : ADBS, 2002, p. 169-201.

LANDRY, Patrice (1999) – *Multilingual subject access project : comparative analysis of titles indexed using LCSH, RAMEAU and SWD/RSWK*. Março 1999. [Em linha]. [Consult. 10 Fev. 2005]. Disponível na WWW: <URL: http://laborix.kub.nl/prj/macspub/2comparative_analysis.pdf>.

LANDRY, Patrice (1997) – *Proposition pour un nouveau catalogue matière à la Bibliothèque nationale suisse : rapport externe*. Suisse : Bibliothèque Nationale Suisse. Service du Catalogage matières, 1997.

LEAL, Filipe (1997) – O fio de Ariane: a organização do conhecimento nas bibliotecas públicas. *Leituras : revista da Biblioteca Nacional* . Lisboa: Biblioteca Nacional. 3ª Série, n.º2 (Out 1997 - Abr 1998), p. 127-140.

LETH, Pia; BERG, Ingrid (2004) – Subject indexing in Sweden : the creating of a national system based on international standards ina country that often wanted to go its own way. In IFLA GENERAL CONFERENCE AND COUNCIL, 70, Buenos Aires, 2004 – [*Proceedings*]. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL:<http://www.ifla.org/IV/ifla70/>>

LOPES, Maria Inês (2000) – Acesso e partilha de informação em rede : interconexão de sistemas com o Z39.50. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas do novo milénio*. Lisboa: Biblioteca Nacional, [2000].

LOPES, Maria Inês (1998) – As bibliotecas e a organização do conhecimento : evolução e perspectivas. *Leituras : revista da Biblioteca Nacional* . Lisboa: Biblioteca Nacional. 3ª Série, n.º2 (Out 1997 - Abr 1998), p. 141 -157.

LOPES, Maria Inês (1995) – PORBASE : Base Nacional de Dados Bibliográficos. *Revista da Biblioteca Nacional*. 2ª série, Vol. 10, n.º1-2 (Jan-Dez 1995), p. 266-271.

LOPES, Maria Inês; BEALL, Julianne, ed. lit. (1999) – *Principles underlying subject heading languages (SHLs)*. München: K. G. Saur, 1999. (UBCIM publications. New Series; 21). ISBN 3-598-11397-8.

LOPEZ ALONSO, Miguel A. (2001) – Integración de teorías para la representación y recuperación del conocimiento. In CONGRESO ISKO ESPAÑA, 5, Alcalá de Henares, 2001 – *La representación y organización del conocimiento : metodologías, modelos y aplicaciones : actas*. Alcalá de Henares: Imagen Documental, 2001. (Organización del conocimiento en sistemas de documentación y información; 5). ISBN 84-8138-435-6. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: <http://www.ugr.es/~isko/>>

LUCARELLI, Anna (2003) – Semantic authority control and the new soggetario. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

LUNAU, Carrol D. (2003) - *Le profil Bath : qu'est-ce que c'est et pourquoi devrais-je m'y intéresser ?* Bibliothèque et Archives du Canada, 2003. [Em linha]. [Consult. 11 Jun. 2003]. Disponível na WWW: <URL: <http://www.nlc-bnc.ca/bath/obj/prof-f-pdf>>

MACS (1999) – *Feasibility study final report*. [Em linha]. [Consult. 14 Fev. 2004]. Disponível na WWW: <URL: <http://laborix.uvt.nl/prj/macs/pub>>

MACS (2002) – *MACS (Multilingual Access to Subjects Project : report for 2000-2001*. [Em linha]. [Consult. 14 Fev. 2004]. Disponível na WWW: <URL: <http://laborix.uvt.nl/prj/macs/pub/MACSreport3.pdf>>

MANIEZ, Jacques (1987) – *Les langages documentaires et classifications*. Paris: Les Éditions d'Organisations, 1987. ISBN 2-7081-0833-6.

MAPLE, Amanda (1997) – *Faceted access : a review of the literature*. [Em linha]. [Consult. 19 Fev. 2005]. Disponível na WWW: <URL: http://theme.music.indiana.edu/tech_s/mla/facacc.rev>

MARTIN, Robert S. (2003) – *Coopération et changements : les centres d'archives, bibliothèques et musées aux États-Unis*. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: http://www.ifla.org/IV/ifla69/papers/066f_trans-Martin.pdf>

MARTÍNEZ ARELLANO, Filiberto Felipe (2004) – *Development of a Spanish subject heading list*. In IFLA GENERAL CONFERENCE AND COUNCIL, 70, Buenos Aires, 2004 – [*Proceedings*]. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL:<http://www.ifla.org/IV/ifla70/>>

MATTELART, Armand (2002) – *História da sociedade da informação*. Lisboa: Bizâncio, 2002. ISBN 972-53-0157-9.

MCGUINNESS, Deborah L.; HARMELEN, Frank van, ed. lit. (2004) – *OWL Web Ontology Language : overview : W3C recommendation 10 February 2004*. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://www.w3c.org/TR/owl-features>>

MENDES, Maria Teresa Pinto ; SIMÕES, Maria da Graça (2001) – Indexação por assuntos: princípios gerais e normas. *Páginas a&b*. Lisboa: Colibri. ISSN 0873-5670. N.º8 (2001), p. 7- 74.

MENON, Bruno (2004) – L'évolution des langages documentaires. [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW: <URL:<http://www.bmenon.net/Evolution%20des%20langages%20documentaires.htm>>

MIETTINEN, Manne (2000) – *Technical standards : XML : [Renardus] Deliverable 2.1 (internal) : technical standards*. [Em linha]. [Consult. 7 Jul. 2000]. Disponível na WWW: <URL:http://www.renardus.org/deliverables/D2_1_Technical_standards.htm>

MILES, A.; MATTHEWS, B.; WILSON, M. (2003) – Making thesauri work for the semantic web. In NKOS WORKSHOP - *Evolving standards*. [Em linha]. [Consult. 16 Mai. 2005]. Disponível na WWW: <http://www.glam.ac.uk/soc/research/hypermedia/NKOS-workshop%20Folder/Miles.ppt>

MILLER, Paul (2000) – I say what I mean, but do I mean what I say?. *Ariadne*. ISSN 1361-3200. N.º 23 (Mar 2000) [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.ariadne.ac.uk/issue23/metadata/intro.html>>

MILLER, Paul (2001) – Interoperability: what is it and why should I want it?. *Ariadne*. ISSN 1361-3200. N.º 29 (Set 2001) [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW: <URL: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>>

MILLER, Paul (1999) – Z39.50 for all. *Ariadne*. ISSN 1361-3200. N.º 21 (Set 1999) [Em linha]. [Consult. 4. Dez. 2000]. Disponível na WWW: <URL: <http://www.ariadne.ac.uk/issue21/z3950/>>

MILLER, Paul; DAWSON, David; PERKINS, John (2001) – Standing on the shoulders of giants: efforts to leverage existing synergies in digital cultural content creation programmes world-wide. *Cultivate Interactive*. N.º 5 (Out. 2001) [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW:

<URL: <http://www.cultivate-int-org/issue5/giants>>

MILSTEAD, Jessica (1999) – *Workshop on electronic thesauri: planning for a standard: report* [Em linha]. [Consult. 13 Fev. 2002]. Disponível na WWW:

<URL: <http://www.niso.org/thes99rprt.html>>

MOEN, William E. (2000) – Interoperability for information access : technical standards and policy considerations. *The journal of academic librarianship*. Vol. 26. n.º 2 (Mar. 2000), p. 129-132.

MOEN, William E. (2003) – Barriers to interoperability : technical and not so technical. In ANNUAL GILS CONFERENCE, 5, Lisle, Illinois, 2003 - [Papers] [Em linha]. [Consult. 11 Jun. 2003]. Disponível na WWW: <URL: http://www.unt.edu/wmoen/presentations/Interoperability_GILS2003_8Apr2003.ppt>

MONTENEGRO, Manuel (1999) – A CDU, monstro pré-histórico das classificações?. *Páginas a&b*. Lisboa: Colibri. ISSN 0873-5670. N.º4 (1999), p. 71-92.

MOREIRA, Lídia Amélia Braga (1980) – *A normalização face às mutações das sociedades*. Lisboa: Minerva, 1980.

MORRIS, Anne (2001) – Knowledge management: opportunities for LIS graduates. In IFLA COUNCIL AND GENERAL CONFERENCE, 67, Boston, 2001 – [Proceedings]. [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla67/papers/015-115ae.pdf>>

NEDOBITY, Wolfgang (1983) – La terminologie et son application à la classification, à l'indexation et à l'analyse. *Révue de l'Unesco pour la science*

de l'information, la bibliothéconomie et l'archivistique. Paris: Unesco. Vol. V, n.º4 (Out-Dez 1983), p. 246-254.

NEUROTH, Heike; KOCH, Traugott (2001) – *Cross-browsing and cross-searching in a distributed network of subject gateways : architecture, data model, and classification*. [Em linha]. [Consult. 14 Out. 2002]. Disponível na WWW: <URL: <http://www.kb.nl/coop/elag/elag2001/www.stk.cz/elag2001/Papers.../HeikeNeuroth.htm>>

NICHOLSON, Dennis (2002) – Subject-based interoperability : issues from the High Level thesaurus (HILT) Project. IFLA COUNCIL AND GENERAL CONFERENCE, 68, Glasgow, 2002 – [*Proceedings*] [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla68/papers/006-122e.pdf>>

NICHOLSON, Dennis [et al...] (2001) - *CAIRNS final report : an embryonic cross-sectorial, cross-domain national networked information service for Scotland?*. Maio 2001. [Em linha]. [Consult. 6 Dez. 2004]. Disponível na WWW: <URL: <http://cairns.lib.gla.ac.uk/cairnsfinal.pdf>>

NICHOLSON, Dennis; MACGREGOR, George (2002) – Learning lessons holistically in the Glasgow digital library. *D-Lib Magazine*. Vol. 8, n.º 7/8 (Jul./Ago. 2002). ISSN 1082-9873. [Em linha]. [Consult. 1 Jul. 2003]. Disponível na WWW: <URL: <http://www.dlib.org/dlib/july02/nicholson/07nicholson.html>>

NICHOLSON, Dennis, WAKE, Susannah (2001) – HILT : subject retrieval in a distributed environment. In IFLA SATTELITE MEETING, Dublin (Ohio), 2001 - *Subject retrieval in a networked environment : proceedings*. München : K.G. Saur, 2003. p. 61-67.

NIEUWENHUYSEN, Paul (2004) - *Z39.50 and related protocols for search and retrieval*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.vub.ac.be/BIBLIO/nieuwenhuysen/courses/chapters/wais.pdf>>

NORTON, M. Jay (1999) – Knowledge discovery in databases. *Library trends*. Illinois: University of Illinois. Graduate School of Library and Information Science. Vol. 48, n.º 1 (Summer 1999), p. 9-21.

NOY, Natalya F. ; McGUINNESS, Deborah L. (2001) – *Ontology development 101 : a guide to creating your first ontology*. Stanford : Stanford University, [2001]. [Em linha]. [Consult. 16 Mar. 2005]. Disponível na WWW: <URL:<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>>

NP 3715. 1989, Documentação – *Método para a análise de documentos, determinação do seu conteúdo e selecção de termos de indexação*. Lisboa: IPQ. 10 p.

NP 4285-3. 2000, Documentação e Informação – *Vocabulário: parte 3: aquisição, identificação e análise de documentos e dados*. Lisboa: IPQ. 23 p.

OLIVEIRA, António Braz de, VASCONCELOS, Manuela; FREIRE, Nuno (2000) - O Projecto MALVINE: manuscritos literários em rede europeia, através do protocolo Z39.50. In 8^{as} JORNADAS PORBASE 2000 – *Sistemas de informação em bibliotecas do novo milénio*. Lisboa: Biblioteca Nacional, 2000.

OTLET, Paul (1934) - *Traité de documentation : le livre sur le livre : théorie et pratique*. Bruxelles : Editiones Mundaneum, 1934.

OYLER, Patricia G. (2001) – Teaching classification in the 21st century. In IFLA COUNCIL AND GENERAL CONFERENCE, 67, Boston, 2001 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2002]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla67/papers/173-142e.pdf>>

PALMER, Sean B. (2001) – *The semantic web: an introduction*. Em linha. [Consult. 21 Mar. 2005]. Disponível em: <URL: <http://informesh.net/2001/swintro/>>.

PEEREBOOM, Marianne (2000) – Technical standards : the Dublin Core Metadata Element Set. In *[Renardus] Deliverable 2.1 (internal) : technical standards*. [Em linha]. [Consult. 7 Jul. 2000]. Disponível na WWW: <URL:http://www.renardus.org/deliverables/D2_1_Technical_standards.htm>

PEEREBOOM, Marianne; DAY, Michael, HUXLEY, Lesly, colab. (2002) – *Documentation of the service organisation (final) and sustainability of central broker service*. [The Hague : KB], 2002. p.10. (Reynard IST-1999-10562; D3.4 / D8.3).

PLACE, Emma (2000) – International collaboration on Internet subject gateways. *IFLA Journal*, N.º 25 (2000) 1. p. 52-56.

PORTUGAL. Biblioteca Nacional (1990) – *CDU – Classificação Decimal Universal: tabela de autoridade*. 2ª ed. Lisboa: BN, 1990. ISBN 972-565-094-8.

PORTUGAL. Biblioteca Nacional (2003) – *Indexação : terminologia e controlo de autoridades (manual)*. Lisboa: BN, 2003. ISBN 972-565-366-7.

PORTUGAL. Biblioteca Nacional. Área de Classificação e Indexação (1998) – *SIPORbase: sistema de indexação em português: manual*. 3ª ed. rev. e aumentada. Lisboa: BN, 1998. ISBN 972-565-154-5.

QIN, Jian; PALING, Stephen (2001) – Converting a controlled vocabulary into an ontology : the caso of GEM. *Information research*. Vol. 6, n.º2 (Jan 2001). [Em linha]. [Consult. 8 Out. 2001]. Disponível na WWW: <URL: <http://information.net/ir/6-2/paper94a.html>>

RANGANATHAN: ahead of his century. [Em linha]. [Consult. 27 Out. 2004]. Disponível na WWW: <URL: <http://www.slais.ubc.ca/courses/libr517/winter2000/Group7/facet.htm>>

RIJSBERGEN, C. J. van; JONES, K. Sparck (1973) – A test for the separation of relevant and non-relevant documents in experimental retrieval collections. *The journal of documentation*. Vol. 29, n.º3 (Set 1973), p. 251-257.

RLG (2005) – *Descriptive metadata guidelines for RLG cultural materials*. Mountain View (California) : RLG, 2005.

ROSTAING, Hervé (2001) – Le Web et ses outils d'orientation. *Bulletin des bibliothèques de France*. T. 46, n.º 1 (2001), p. 68-77.

ROWLEY, Jennifer E. (1987) – *Organising knowledge : an introduction to information retrieval*. Aldershot, Hants; Brookfield, Vermont: Gower, 1987. ISBN 0-566-03486-7(Pbk)

RUSSEL, Rosemary; DAY, Michael (2001) – *HILT : High-Level Thesaurus . automated and manual approaches to the provision of thesauri and subject vocabularies*. Bath: UKOLN, 2001. (Hilt-interface-10.doc; final version). [Em linha]. [Consult. 12 Jan. 2004]. Disponível na WWW: <URL: <http://hilt.cdlr.strath.ac.uk/Reports/Documents/hilt-interface-10.doc>>

SAINT PIERRE, Margaret [et al...] (1994) - *WAIS over Z39.50-1988*. [Em linha]. [Consult. 22 Mar. 2005]. Disponível na WWW: <URL: <http://www.fags.org/rfcs/rfc1625.html>>

SALTON, G.; YANG, C. S. (1973) – On the specification of term values in automatic indexing. *The journal of documentation*. Vol. 29, n.º4 (Dez 1973), p. 351-372.

SANTOS, Lúcia (2000) - A cooperação na PORBASE : dados estatísticos. In JORNADAS PORBASE, 8, Lisboa, 2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000].

SANTOS, Luísa (2000) – Relançamento do Projecto CLIP : Compatibilização de Linguagens de Indexação em Português. In Jornadas PORBASE, 8, Lisboa,

2000 - *Sistemas de informação em bibliotecas no novo milénio*. Lisboa: Biblioteca Nacional, [2000].

SLAVIC, Aida (2003) - UDC implementation: from library shelves to a structured indexing language. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla69/papers/032e-Slavic.pdf>>

SOUSA, Manuel de, coord. (2001) – *Dicionário de termos informáticos : milénio : inglês / português, português / inglês, internet / redes, sites 2001*. Mem Martins : SporPress, 2001.

SRW – *Search/Retrieve web Service*. [Em linha]. [Consult. 31 Jan. 2005]. Disponível na WWW: <URL: <http://www.loc.gov/z3950/agency/zing/srw/background.html>>

STIFF, Matthew; HOLM, Stuart – *mda : museums and the High Level Thesaurus Concept. HILT Phase I : reports*. [Em linha]. [Consult. 18 Fev. 2005]. Disponível na WWW: <URL: <http://hilt.cdlr.strath.ac.uk/Reports/museumsreport.html>>.

STOKLASOVA, Bohdana (2003) – Short survey of subject gateways (SG) activity. In IFLA GENERAL CONFERENCE AND COUNCIL, 69, Berlin, 2003 – [*Proceedings*]. [Em linha]. [Consult. 2 Out. 2003]. Disponível na WWW: <URL: <http://www.ifla.org/IV/ifla69/papers/152e-Stoklasova.pdf>>

SVENONIUS, Elaine (2004) – The epistemological foundations of knowledge representation. *Library trends*. Vol. 52, n.º 3 (Winter 2004), p. 571-587.

TARTAGLIA, Stefano (2003) – Controllo d'autorità e linguaggi di indicizzazione per soggetto. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

TAYLOR, Mike (2003) – *Zthes profile for SRW, version 0.5*. 12 Dez. 2003. [Em linha]. [Consult. 31 Jan. 2005]. Disponível na WWW: <URL: <http://zthes.z3950.org/srw/current.html>>.

TENNIS, Joseph T. (2003) – Data collection for controlled vocabulary interoperability – Dublin Core audience element. *Bulletin of the American Society for Information Science and Technology*. Vol. 29, n.º2 (Dez-Jan 2003), p. 20-23.

TILLET, Barbara B. (2003) – Authority control : stato dell'arte e nuove prospettive. In INTERNATIONAL CONFERENCE, Florence, 2003 – *Authority control : definition and international experiences*. [Em linha]. [Consult. 24 Nov. 2004]. Disponível na WWW: <URL: <http://www.sba.unifi.it/ac/>>

TILLET, Barbara B. (2002) – A Virtual International Authority File (VIAF). In WORKSHOP ON AUTHORITY CONTROL AMONG CHINESE, KOREAN AND JAPANESE LANGUAGES, 3, Karuizawa, Tokyo, Kyoto, 2002. [Em linha]. [Consult. 12 Jan. 2004]. Disponível na WWW: <URL: <http://www.nii.ac.jp/publications/CJK-WS3/cjk3-08a.pdf>>.

TOUS LES SAVOIRS DU MONDE : liste des textes fondateurs : quelques encyclopédies et textes fondateurs de la pensée occidentale, avant l'explosion des savoirs. [Em linha]. [Consult. 28 Fev. 2005]. Disponível na WWW: <URL: <http://classes.bnf.fr/dossism/textfond.htm#XVle>>.

UNESCO (1971) – *UNISIST : étude sur la réalisation d'un système mondial d'information scientifique*. Paris: UNESCO, 1971.

VIZINE-GOETZ, D. (2003) – Terminology services. In NKOS WORKSHOP - *Evolving standards*. [Em linha]. [Consult. 16 Mar. 2005]. Disponível na WWW: <URL:<http://www.glam.ac.uk/soc/research/hypermedia/NKOS-workshop%20Folder/vizine-goetz.ppt>>

VIZINE-GOETZ, Diane *et al.* (2004) - Vocabulary Mapping for Terminology Services. *Journal of Digital Information*. Vol. 4, N.º 4, Artigo n.º 272 (2004-03-11). [Em linha]. [Consult. 16 Mar. 2005]. Disponível na WWW: <URL: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/>>

W3C (2004) – *OWL Web Ontology Language : use cases and requirements* . W3C recommendation 10 February 2004. [Em linha]. [Consult. 17 Nov. 2004]. Disponível na WWW: <URL: www.w3.org/TR/webont-req/>

WC3. Semantic Web Activity (2005) – *Semantic Web*. [Em linha]. [Consult. 21 Mar. 2005]. Disponível na WWW: <URL: <http://www.w3.org/2001/sw/>>

WC3. Semantic Web Activity (2004) – *Resource Description Framework (RDF)*. [Em linha]. [Consult. 21 Mar. 2005]. Disponível na WWW: <URL: <http://www.w3.org/RDF/>>

WAKE, Susannah; NICHOLSON, Dennis (2001) – HILT – High-Level Thesaurus Project : building consensus for interoperable subject access across communities. *D-Lib Magazine*. Vol. 7, n.º 9 (Set. 2001). ISSN 1082-9873. [Em linha]. [Consult. 21 Mai. 2003]. Disponível na WWW: <URL: <http://www.dlib.org/dlib/september01/wake/09wake.html>>

WANG, Jun (2003) – A Knowledge network constructed by integrating classification, thesaurus and metadata in digital library. *Bulletin of the American Society for Information Science and Technology*. Vol. 29, n.º2 (Dez-Jan 2003), p. 24-28.

Zthes : a profile for thesaurus navigation in Z39.50 and SRW. 17th December 2004. [Em linha]. [Consult. 31 Jan. 2005]. Disponível na WWW: <URL: <http://zthes.z3950.org/>>

ANEXOS

Registos de autoridade (Unimarc/Autoridades)

Registo de autoridade de pessoa física:

000		00301cx 022001093 45
001		1118414
152		\$aRPC
200	1	\$aGedeão,\$bAntónio,\$cpseud.
305		\$aPara a obra científica, ver também\$bCarvalho, Rómulo de, 1906-1997
500	1	\$aCarvalho,\$bRómulo de,\$f1906-1997
801	0	\$aPT\$bBN
830		\$9PT\$aPoeta. Romancista

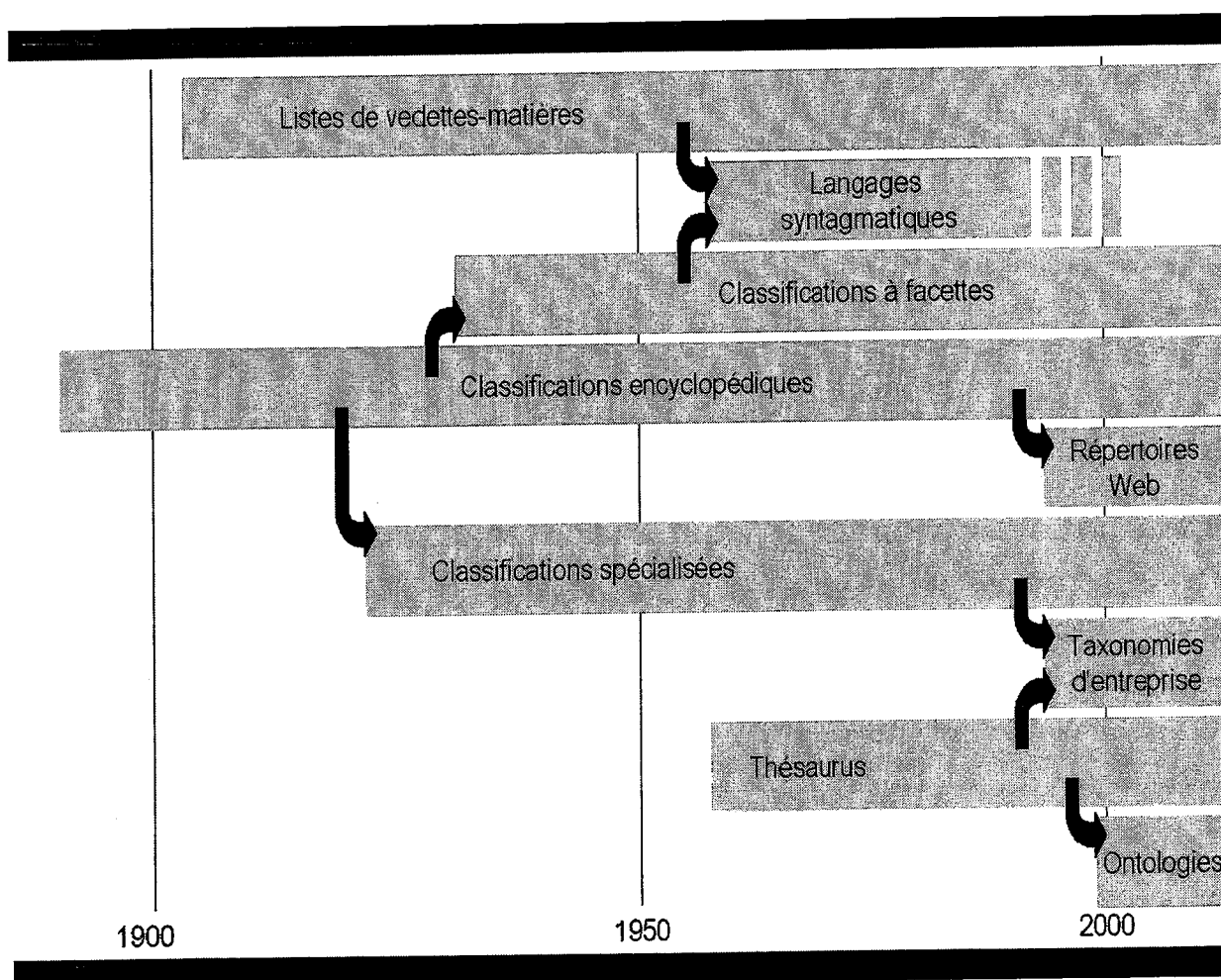
Registo de autoridade de colectividade:

000		00679cx 2200241 45
001		10631
095		\$aPTBN00007542
100		\$a19900620apory0103 ba
152		\$aRPC\$bSIPOR
210	0 2	\$aFundação Calouste Gulbenkian
400	0	\$aGulbenkian
410	0 2	\$aFCG
410	0 2	\$aFundação Gulbenkian
410	0 2	\$aCalouste Gulbenkian Foundation
410	0 2	\$aFondation Calouste Gulbenkian
550		\$aInstituições de beneficiência
550		\$aInstituições artísticas
550		\$aInstituições de educação
550		\$aInstituições científicas
675		\$a061.22(469) FCG\$vBN\$zpor
801	0	\$aPT\$bBN\$c19900620
810		\$aOrdenações Afonsinas
102		\$aPT

Registo de autoridade de assunto (nome comum):

000		00766nx 2200193 45
001		9556
095		\$aPTBN00007141
100		\$a19900612apory0103 ba
152		\$bCLIPBA
250		\$aArquitectura
305	1	\$aVeja também a subdivisão\$bArquitectura sob nomes de localidades.
330	1	\$aAplica-se a obras sobre a arte de construir edificios. A arquitectura de um determinado país ou povo deve ser representada pelo termo arquitectura modificado pelo adjectivo de nacionalidade ou povo adequado.
550		\$aArquitectura civil
550		\$aArquitectura militar
550		\$aArquitectura habitacional
550		\$aPatrimónio arquitectónico
550		\$aProjectos de arquitectura
550		\$aEspaço arquitectónico
675		\$a72\$vBN\$zpor
801	0	\$aPT\$bBN\$c19900612

Quadro genealógico das linguagens documentais



In: MENON, Bruno (2004) – *L'évolution des langages documentaires*. [Em linha]. [Consult. 9 Nov. 2004]. Disponível na WWW:
<URL:<http://www.bmenon.net/Evolution%20des%20langages%20documentaires.htm>>
[Artigo preparado para a revista *Les Dossiers de l'ingénierie éducative* n° 49, décembre 2004]

Construir uma ontologia¹

Preparar uma chávena de chá

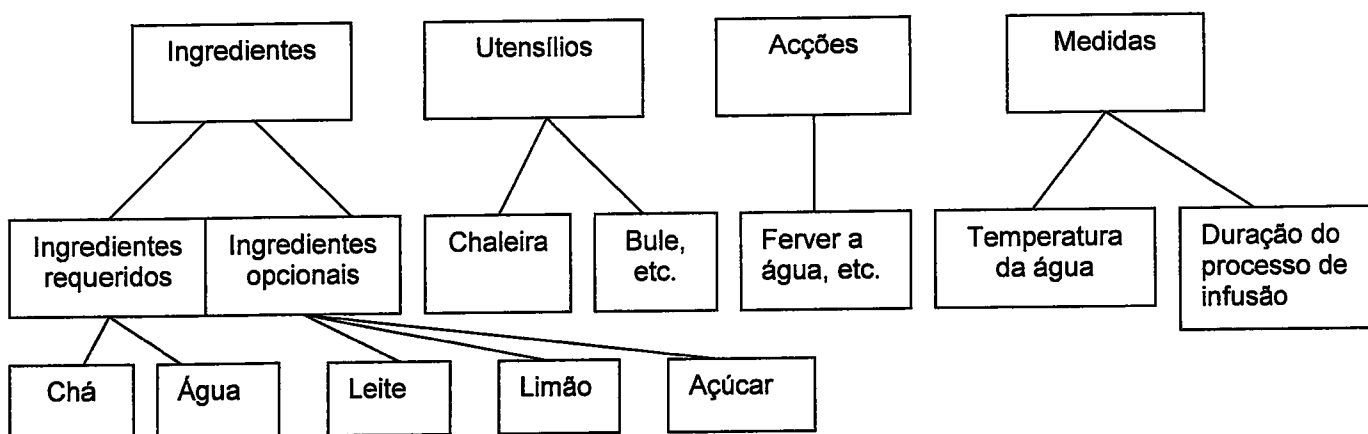
1. Estabelecer a finalidade e objectivo da nossa ontologia

- Descrever o processo de fazer chá, para o automatizar
- Avaliar a qualidade de uma chávena de chá específica
- Avaliar as aptidões de um aprendiz de “fazedor de chá”
- Escolher o melhor tipo de chá para ser servido num tempo determinado
- Permitir às casas de chá a partilha de diferentes dados estruturados...

2. Identificar as entidades que são específicas do domínio

- chá, água, leite ou limão, açúcar...
- chaleira, bule, passador de chá, chávena, colher...
- ferver a água, deitar a água fervente no bule, deitar a água fora do bule, ferver lentamente alguma água fresca...
- temperatura da água, duração do processo de infusão...

3. Organizar as entidades numa hierarquia



4. Definir as entidades

- O que é um “ingrediente”?
- O que é um “ingrediente requerido”?
- O que é um “ingrediente opcional”?
- O que é que se entende por “chá”?

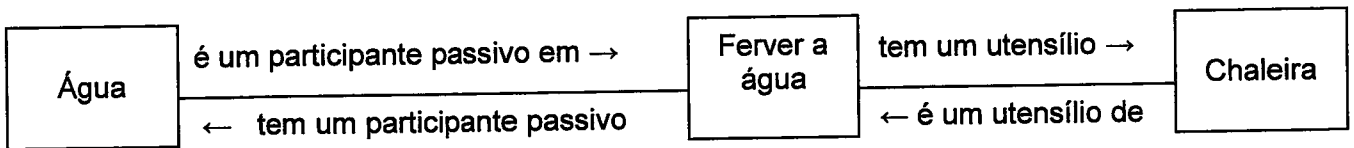
¹ Adaptado de CROFTS, Nicolas, LE BOEUF, Patrick, ARTUR, Odile (2002) – ELAG Presentation : Ontologies. *Semantic web and libraries* / European Library Automation Group 2002; 26th Library systems seminar, Rome, 2002. [Em linha]. [Consult. 19 Set. 2002]. Disponível na WWW: <URL: <http://www.kb.nl/coop/elag/elag2002/www.ifnet.it/elag2002/papers/pap9.html>>.

5. Propriedades das entidades

Para caracterizar o açúcar, por exemplo, podemos precisar das seguintes propriedades, cada uma das quais pode ter um ou vários valores.

Origem	Cana, beterraba, ácer
Cor	Branco, escuro
Forma	Granulado, em pó, em cubos...

6. Identificar as relações

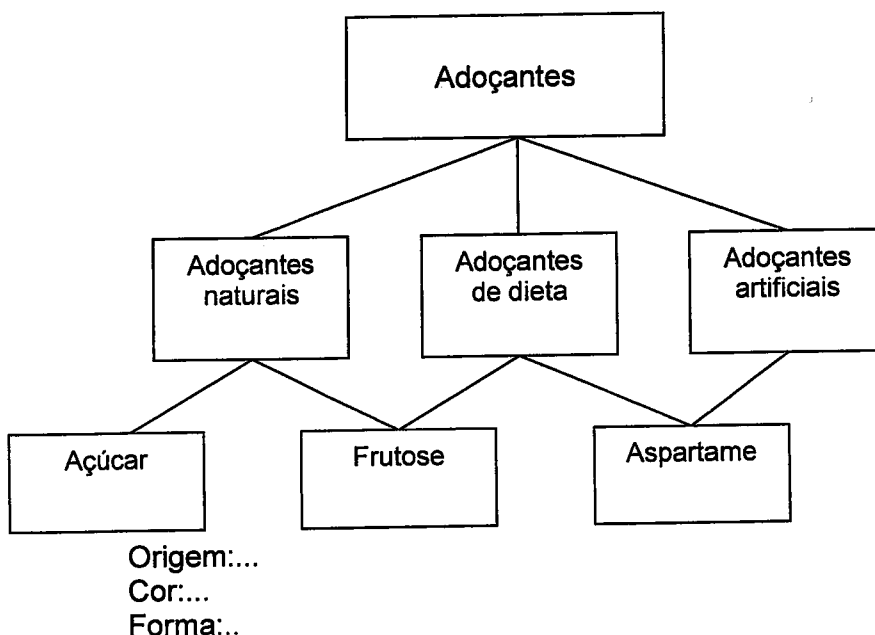


7. Descrever e definir relações

Naturalmente as relações têm de ser *descritas* e *definidas* do mesmo modo que as entidades, pelas mesmas razões.

8. Refinamentos e extensões

Neste ponto pode parecer que o trabalho está acabado – contudo, trata-se apenas do início, agora para a tarefa de refinar, expandir e desenvolver o começo da ontologia – comparando-a com outras fontes de informação e testando-a para assegurar que efectivamente cobre o domínio pretendido. As aplicações específicas podem requerer níveis adicionais de detalhe, com os quais a ontologia terá de lidar. No nosso exemplo, o açúcar é a única forma de adoçante que foi reconhecida, mas existem outras. Expandir a ontologia para incluir outros produtos implica a criação de uma nova entidade “adoçante”, que pode incluir açúcar, sacarina, mel, etc.



Elementos e qualificadores de metadados que fazem parte do perfil de aplicação Renardus (Maio 2001)

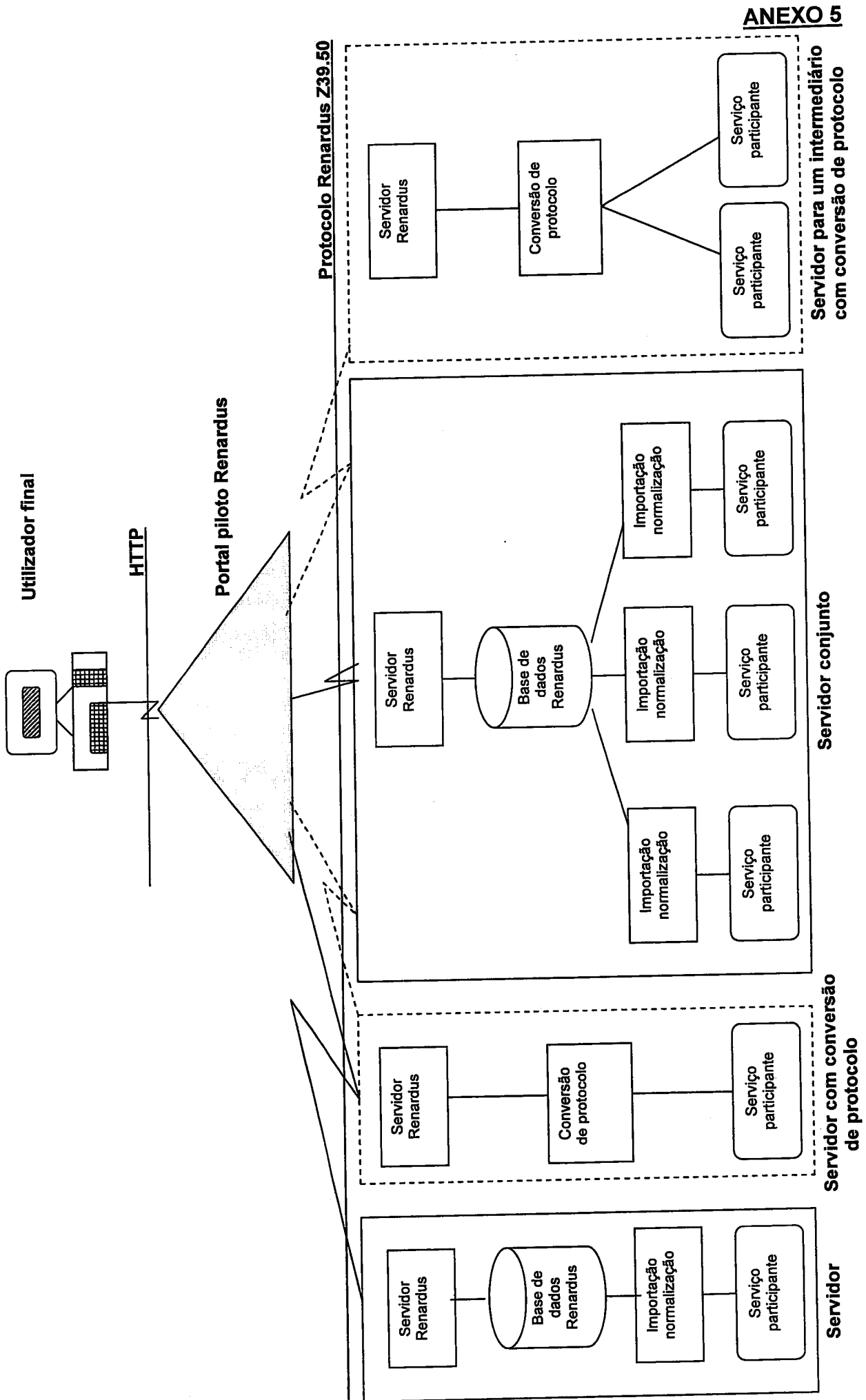
Metadata Element	Obligation	Repeatable	LQ	Namespace	Comments
dc:title	M	NR	yes	DCMES	-
dc:title.alternative	O	R	yes	DCMES Qualifiers	-
dc:creator	R	R	no	DCMES	-
dc:creator (R Qualifiers)	R	R	no	RMES Qualifiers	Creator(s) are person(s) which are responsible for the intellectual content of the document(s), e.g. webmasters are no creators. If this field is applicable it is strongly recommended to provide the creator. For Renardus normalization process it is strongly recommended that last name and first name are clearly distinguishable.
dc:description	M	R	yes	DCMES	For cross-search reasons the field description must contain free text.
dc:subject	M	R	yes	DCMES	In the prototype system there will be no further distinction between the several kinds of subject (keywords, classification system) and the provision of keywords is strongly recommended. In the final system the provision of keywords is required.
dc:subject	R	R	yes	DCMES Qualifiers	-
dc:subject (R Qualifiers)	M	R	yes	RMES Qualifiers	All other encoding schemes used by the partners, see Acronym & Abbreviation list at http://renardus.sub.uni-goettingen.de/renap/racr.html .
dc:subject.Ren-DDC	M	R	no	RMES Qualifiers	DDC 21: adapted DDC version for cross-browsing purpose. Only captions and not notations will be displayed.
dc:identifier	M	R	no	DCMES Qualifiers	In the prototype system no distinction will be made between resource URL, mirrored, copied resource URL(s) and URL(s) for archive reasons.
dc:identifier (R Qualifiers)	O	R	yes	RMES Qualifiers	Renardus refinements for translated sites and/or mirrored, copied sites, will be realized in the final version.
dc:language	R	R	no	DCMES Qualifiers	The language code is the ISO 639-2, three letter code. A mapping between the two letter (ISO 639-1) and three letter language code will be found on the LoC site: http://lweb.loc.gov/standards/iso639-2/englangn.html

ANEXO 4 (cont.)

Metadata Element	Obligation	Repeatable	LQ	Namespace	Comments
dc:type	R	R	no	DCMES	Subject gateways should provide their original types without encoding scheme.
dc:type.DCT1	R	R	no	DCMES Qualifiers	A mapping from partners' type list to DCMI Type Vocabulary (DCT1) is strongly recommended.
Country	R	NR	no	RMES Qualifiers	3166-1 (two letter code)
Full Record URL	R	NR	no	RMES Qualifiers	A URL that leads to a detailed display of each record at the originating service site.
SBIG ID	M	NR	no	RMES Qualifiers	A stable unique acronym also well defined in the Collection Level Description.

In: NEUROTH, Heike; KOCH, Traugott (2001) – *Cross-browsing and cross-searching in a distributed network of subject gateways : architecture, data model, and classification*. [Em linha]. [Consult. 14 Out. 2002]. Disponível na WWW: <URL: <http://www.kb.nl/coop/elag/elag2001/www.stk.cz/elag2001/Papers.../HeikeNeuroth.htm>>

Arquitectura do Projecto Piloto RENARDUS



ANEXO 5

In: ARDÓ, Anders [et al.] – Specification of Renardus Pilot broker system : Renardus deliverable D2.3 version 1.3. 2001 (tradução nossa).

Ferramenta de mapeamento RENARDUS

▲ ZU+000 Mining General
ddc:

- ZU+060 Mining Districts
ddc:
- ZU+150 Mining Law
- ZU+180 Mineral Economics
ddc.
- ZU+200 Education, Jobs, Future
- ZU+400 Mining History
ddc:
- ZUB+000 Mine Surveying, Prospecting
- ZUE+000 Geomechanics, Rock Mechanics
- ✓ ZUG+000 Mining Technology
ddc:
- ✓ ZUS+000 Mining Branches
ddc:

--- Main Classes

--- 6.. Technology
GOK

- 62- Engineering
 - ✓ 620 Engineering and allied operations
 - ✓ 621 Applied physics
 - ▼ 622 Mining and related operations
GOK
 - ✓ 623 Military and nautical engineering
 - ✓ 624 Civil engineering
 - ✓ 625 Engineering of railroads and roads
626
 - ✓ 627 Hydraulic engineering
 - ✓ 628 Sanitary and municipal engineering
 - ✓ 629 Other branches of engineering

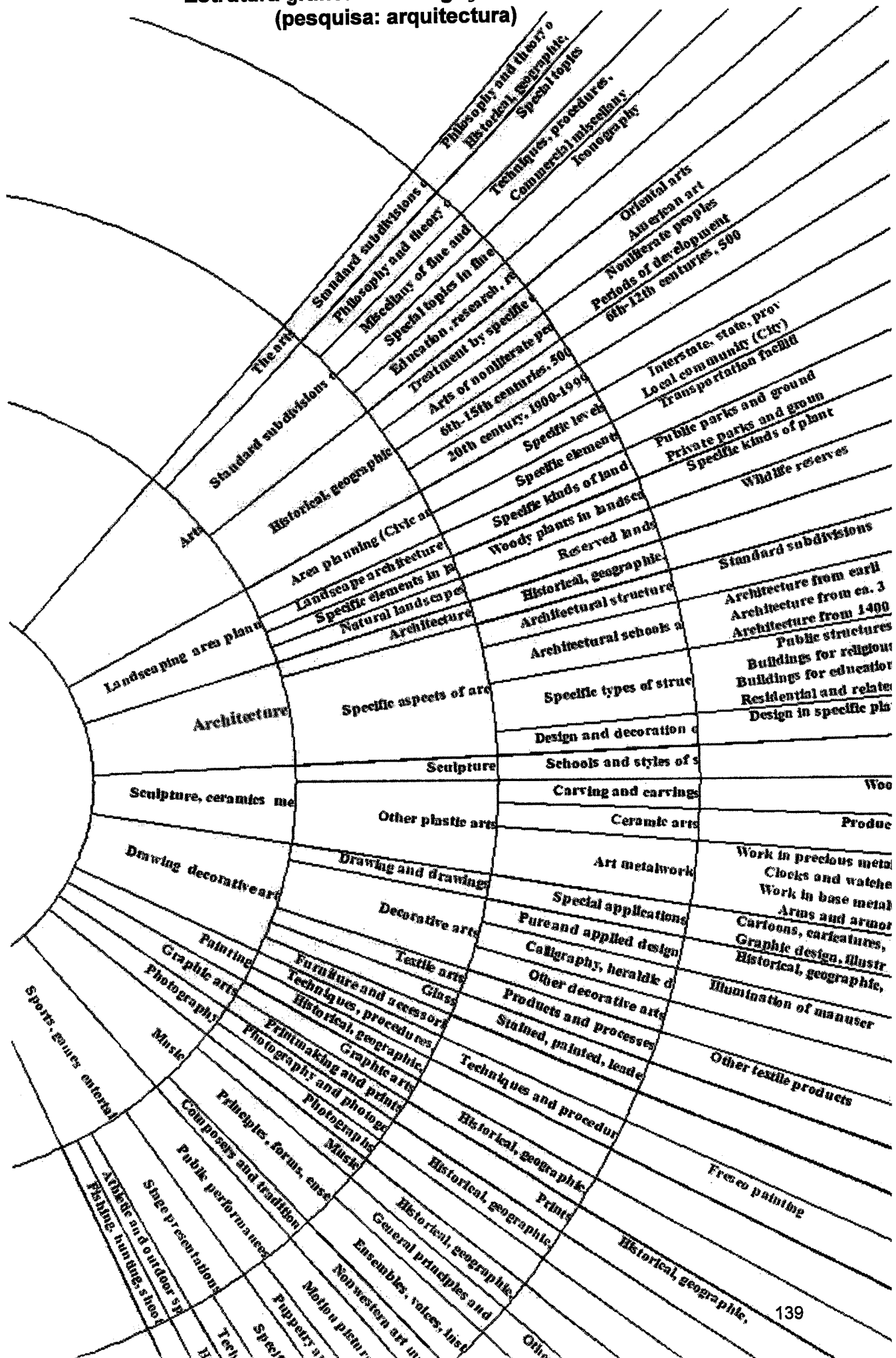
existing relation (one-sided →)

GOK	relation	ddc	relevance	remark
ZU+000	LNE > →	622	→	
cap 1	Mining General			
cap 2	Mining and related operations			
url	http://www.Geo-Guide.de/cgi-bin/ssgfi/anzeige.pl?db=geo			

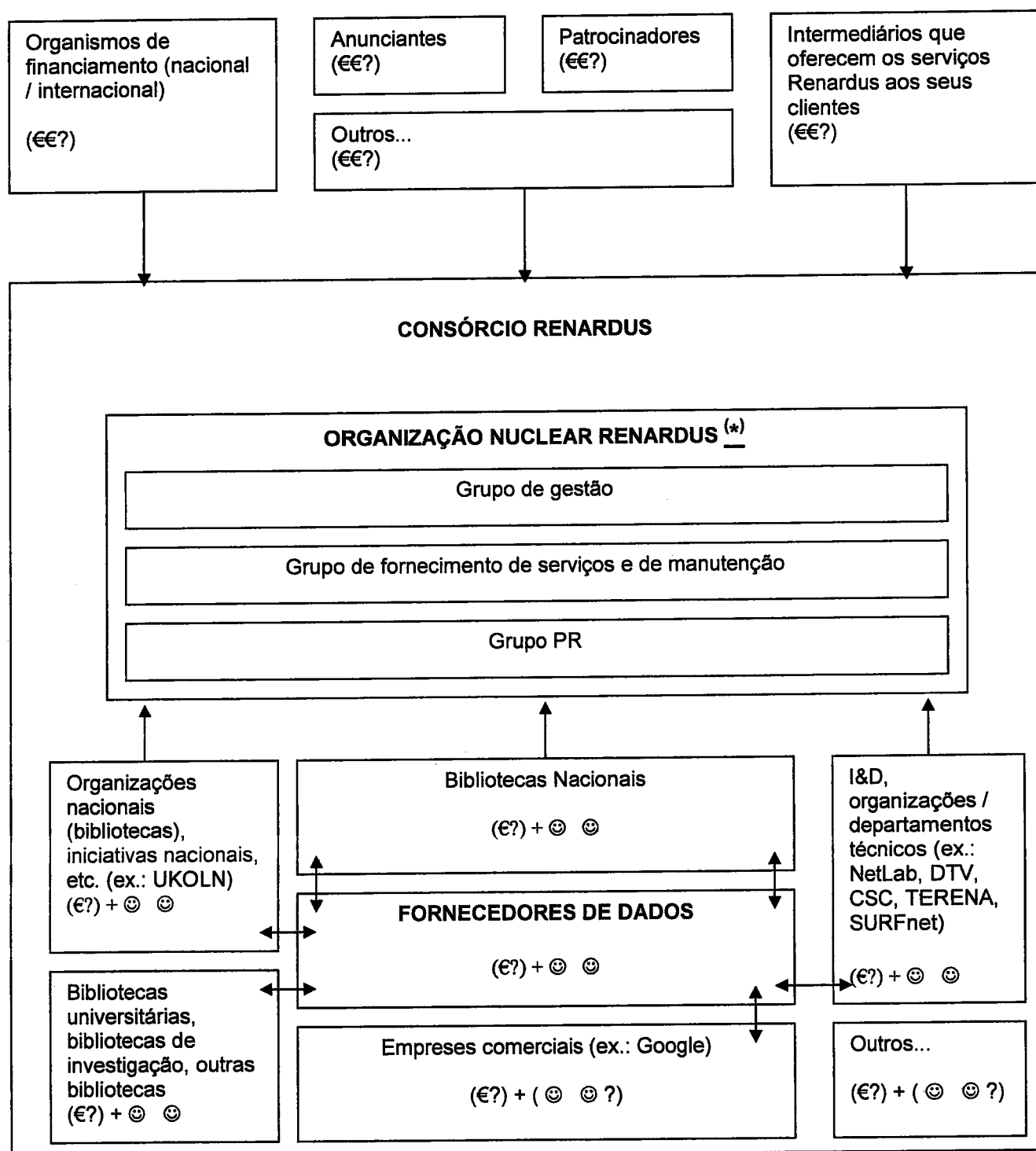
Fig. 1: The Renardus mapping tool

KOCH, Traugott; NEUROTH, Heike; DAY, Michael (2001) – Renardus : cross-browsing European subject gateways via a common classification system (DDC). In IFLA SATTELITE MEETING, Dublin (OH), 2001 - *Subject retrieval in a networked environment : proceedings..* München : K.G. Saur, 2003. p.28.

Estrutura gráfica de navegação Renardus
(pesquisa: arquitectura)



Modelo organizacional para o serviço intermediário RENARDUS



€ - Indica financiamento

☺ - Indica possíveis contributos na forma de recursos humanos (regulares) na própria organização/serviço

☺ - Indica possíveis participantes no Grupo de Gestão, no Grupo de fornecimento de serviços e de manutenção, ou no Grupo PR

(*) Atribuições e competências dos grupos:

Grupo de gestão:

- Administração do consórcio
- Definição da política do serviço / estratégias de negócio
- Provisão de serviços / inclusão de novos portais
- Gestão financeira
- Gestão legal e de direitos

Grupo de fornecimento de serviços e de manutenção:

- Manutenção da infraestrutura técnica e do apoio técnico
- Inovação e desenvolvimento
- Manutenção do modelo de dados
- Manutenção da informação do mapeamento

Grupo PR:

- Manutenção das páginas informativas no sítio *web* do serviço e nas páginas mais restritas
- Promoção e *marketing* do serviço / difusão
- Apoio ao utilizador / *Helpdesk* (utilizador final)
- Apoio aos fornecedores de dados, actuais e potenciais
- Recolha de pareceres dos utilizadores e estatísticas
- Facilitação da comunicação

In: PEEREBOOM, Marianne; DAY, Michael, HUXLEY, Lesly, colab. – *Documentation of the service organisation (final) and sustainability of central broker service*. [The Hague : KB], 2002. p. 10. (Reynard IST-1999-10562; D3.4 / D8.3). [Tradução nossa]

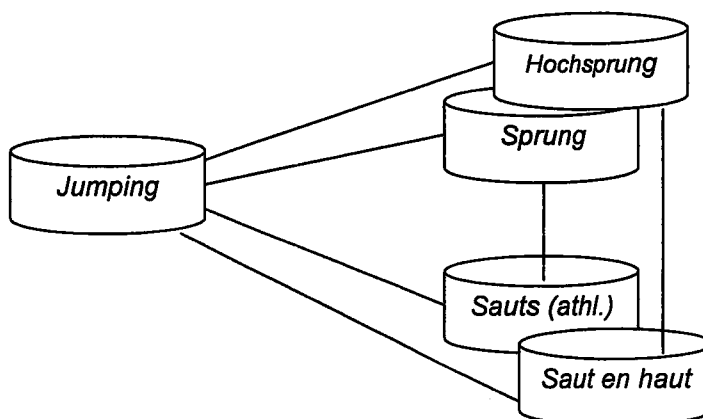
Projecto MACS

Modelos de ligação*

1. Múltiplas correspondências

Ligação um para dois Ex.: *Jumping*

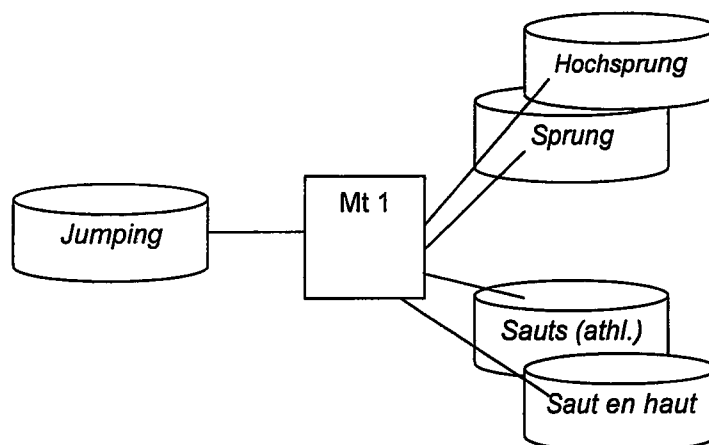
A. Ligação directa entre as três listas de autoridade



No caso de se criar um segundo cabeçalho dos LCSH, duas ligações para serem modificadas pela British Library, com repercussões na SWD e no RAMEAU.

B. Ligação através do "metatesauro":

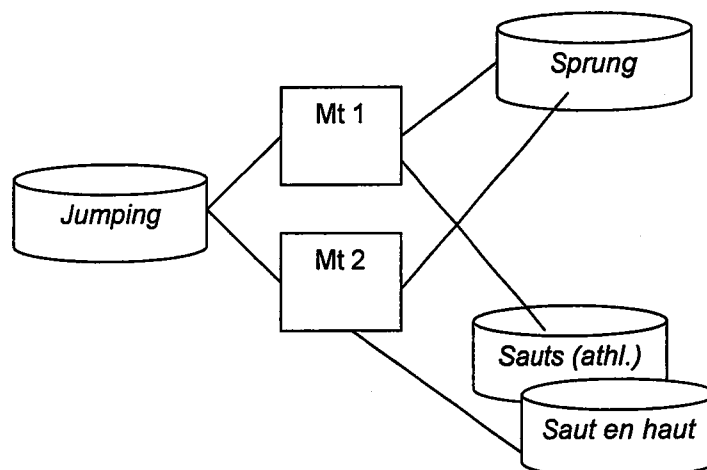
B.1 Com as LCSH como "pivot":



Esta solução não parece boa:

- Perdem-se as equivalências exactas entre as SWD e o RAMEAU;
- No caso de se criar uma nova LCSH, necessidade de criar um novo "Mt" (Mt.= registo do "metatesauro") e de modificar 3 ligações (uma por cada parceiro).

B.2 Sem "pivot":

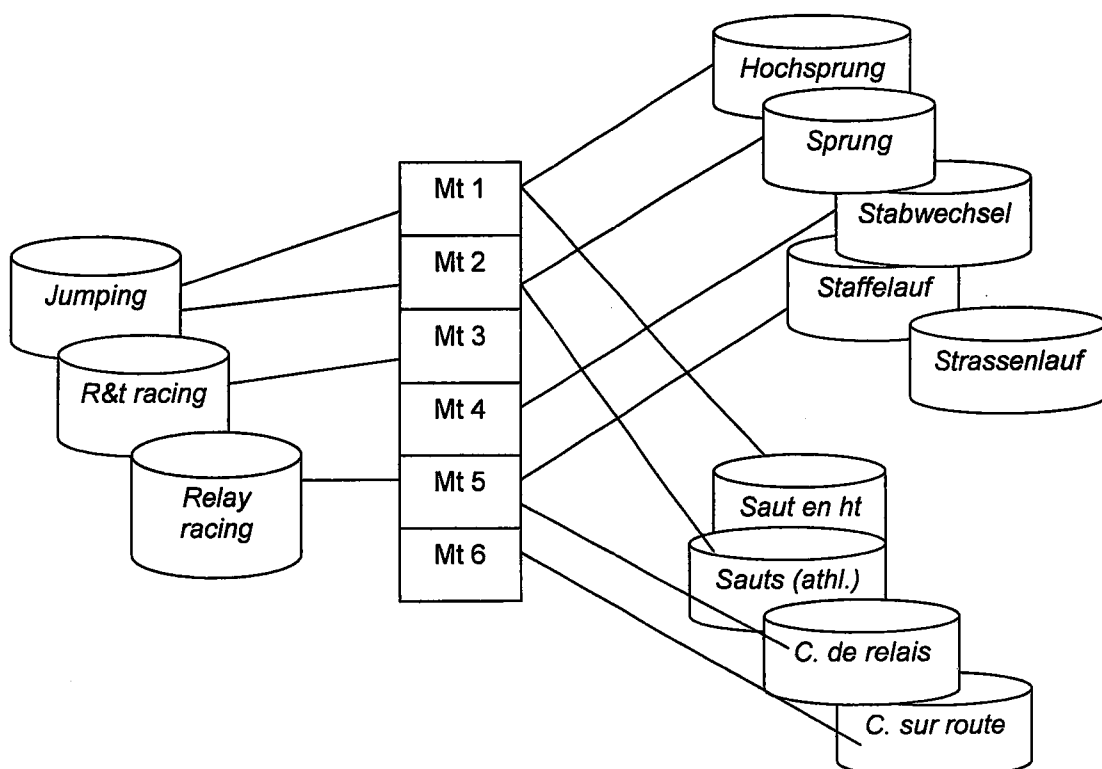


- A melhor solução, com um Mt (= registo do "metatesauro") para cada entrada identificada: as equivalências SWD – RAMEAU são respeitadas.
- No caso de se criar uma nova LCSH, problemas de gestão muito ligeiros: uma ligação a ser feita pela British Library, sem repercussão a nenhum nível (Mt, SWD ou RAMEAU).

2. Modelo global

A. "Metatesauro" com tantos registos quantas as entradas identificadas:

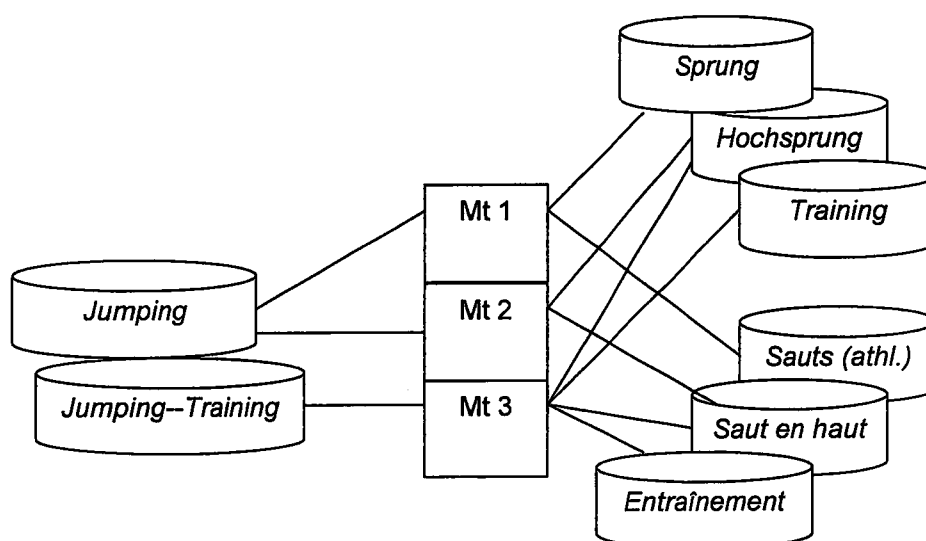
Cada entrada identificada (trilingue, bilingue ou apenas monolingue) está ligada a um registo específico no Metatesauro externo (Mt...).



B. "Metatesauro", possíveis soluções para correspondências múltiplas:

Ligação um para dois (exemplo: *Jumping*) e problemas de encadeamento (ex.: *Jumping – Training = Hochsprung + Training = Saut en hauteur + Entraînement*)

Registo do Metatesauro (Mt...) pode ligar 2 cabeçalhos em caso de equivalência de "cabeçalho para encadeamento"



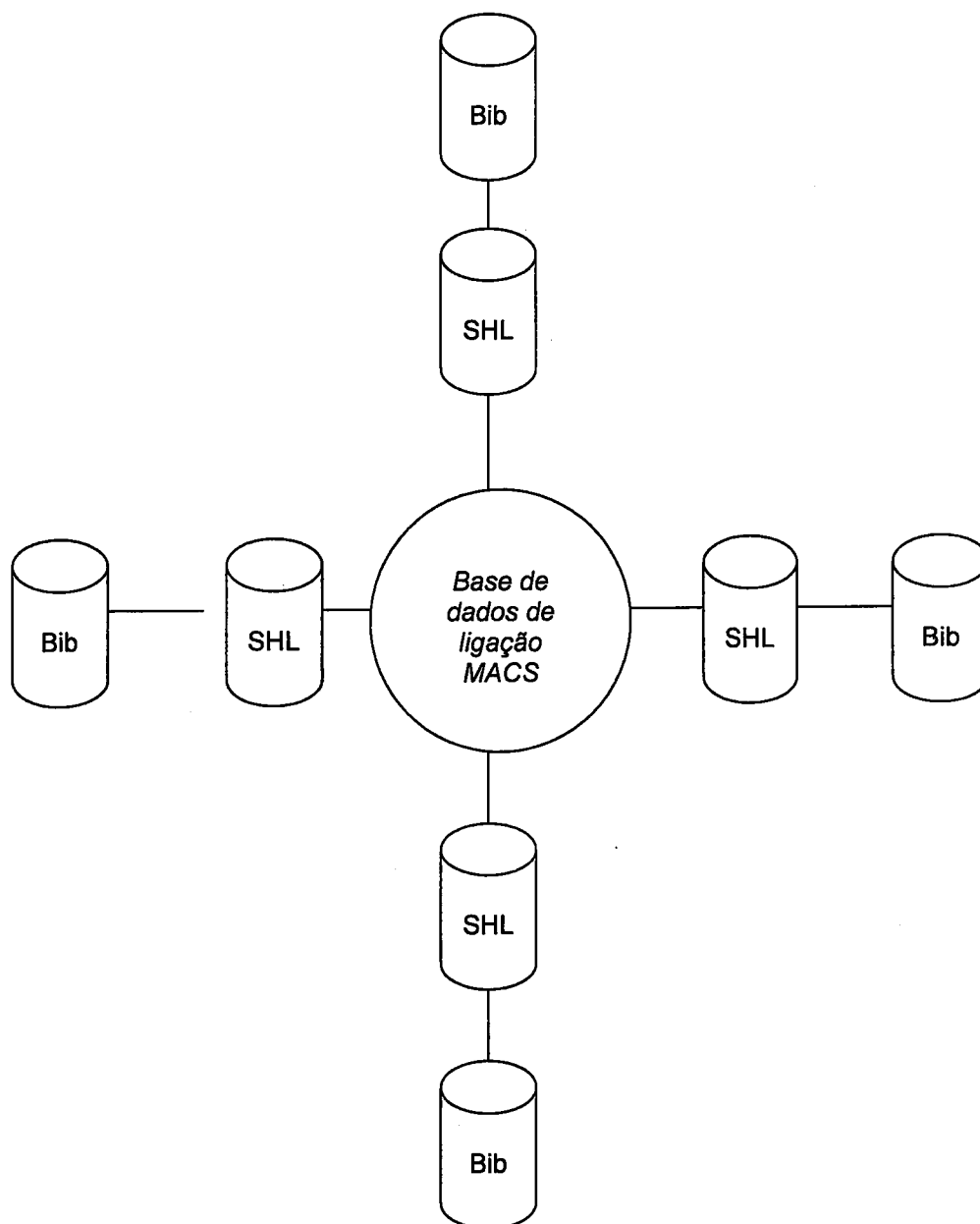
NB: do ponto de vista de British Library:

Jumping = Sprung "ou" Hochsprung = Saut (athlétisme) "ou" Saut en hauteur

Jumping – Training = Hochsprung "e" Training = Saut en hauteur "e" Entraînement

* Fonte : MACS - *Feasibility study final report* (MACS Documentation).[Em linha]. [Consul. 14 Fev. 2005]. Disponível na <URL: <http://laborix.uvt.nl/prj/macspub>>.

Arquitectura do sistema MACS



SHL – *Subject Headings Language*

In: HOPPENBROUWERS, Jeroen (2001) – *MACS : Multilingual Access to Subjects : architecture of the MACS system*. Denmark: Index Data Aps; Tilburg : University, 2001. p. 4.

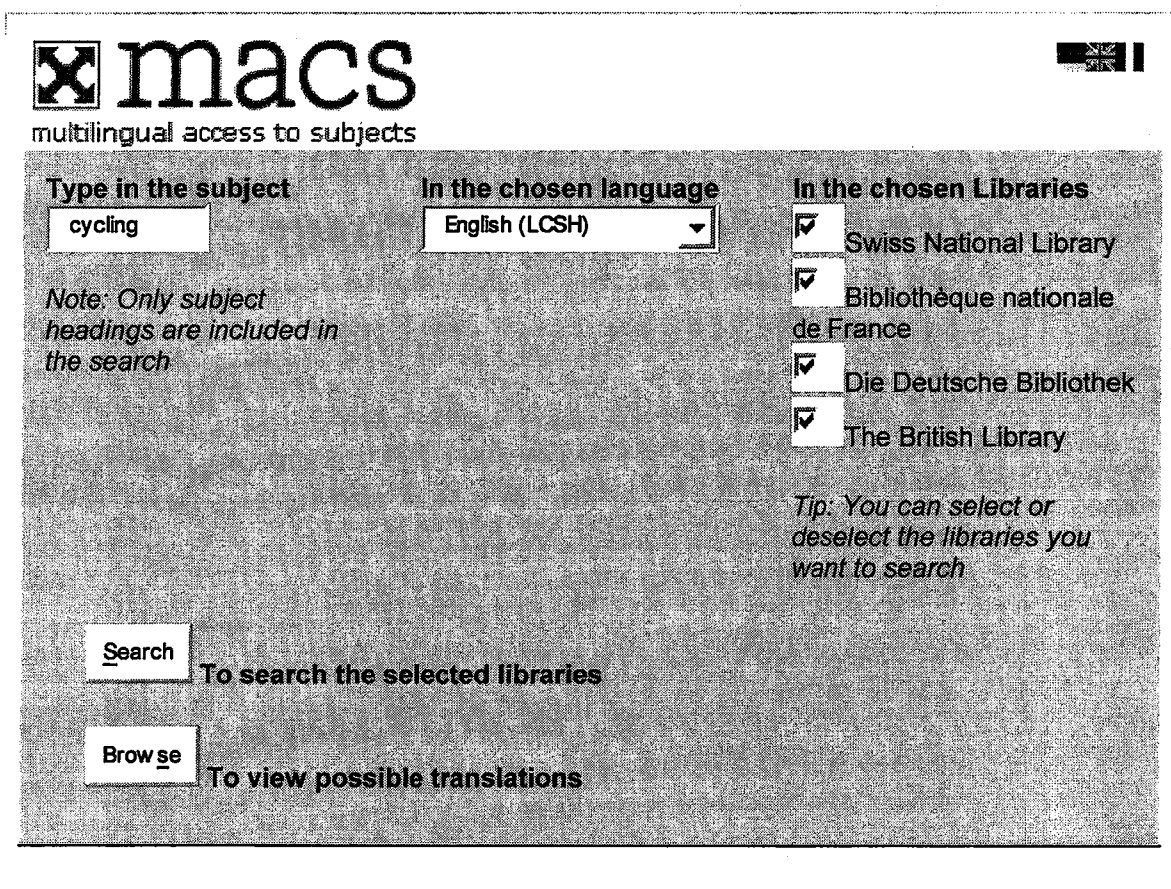
Projecto MACS**Exemplos de equivalência estabelecida entre as linguagens LCSH, SWD e RAMEAU**

English (LCSH)	Deutsch (SWD)	Français (RAMEAU)
Buildings	Gebäude or Bauwerk	Constructions or Constructions
Church buildings	Kirchenbau	Églises
Historic buildings	Baudenkmal	Monuments historiques

English (LCSH)	Deutsch (SWD)	Français (RAMEAU)
Cycling	Radsport or Radfahren	Cyclisme or Cyclisme
Cycling accidents	Sportverletzung and Radsport	Cyclistes – Lésions et blessures
Cycling – Law and legislation	Sportrecht and Radsport	Droit and Cyclisme
Cycling – Training	Training and Radsport	Entraînement and Cyclisme

Fonte : <URL : <http://muffin.indexdata.dk/macs/>> [Consult. 2 Dez. 2004].

Exemplo de pesquisa no protótipo MACS



macs
multilingual access to subjects

Type in the subject
cycling

In the chosen language
English (LCSH)

In the chosen Libraries

- Swiss National Library
- Bibliothèque nationale de France
- Die Deutsche Bibliothek
- The British Library

Note: Only subject headings are included in the search

Tip: You can select or deselect the libraries you want to search

Search To search the selected libraries

Browse To view possible translations



Your search results in the libraries...

[Home](#) | [Browse translations](#)

British Library: 30 hits.

1. [The complete book of mountain biking; Brant Richards and Steve Worland;](#)
2. [The mountain bike book; David Leslie; photographs by Tim Woodcock;](#)
3. [Mountain bike racing; Tim Gould and Simon Burney;](#)
4. [Mountain biking; Brant Richards;](#)
5. [Learn mountain biking in a weekend;](#)
6. [Fat tire; a celebration of the mountain bike; Amici Design;](#)
7. [Pro mountain biker; the complete manual of mountain biking - bikes accessories and techniques; Jeremy Evans and Brant Richards;](#)
8. [Mountain biking; Paul Skilbeck;](#)
9. [The ultimate mountain bike book;](#)
10. [Advanced mountain biking;](#)

[Show more records](#)

Bibliothèque nationale de France: 66 hits.

1. [La Consultation médicale du cycliste; D J.-P. de Mondenard;](#)
2. [Le Peloton de tÃete; Jean-Baptiste Bellone;](#)
3. [Les Cyclistes en questions; D J.-P. de Mondenard; dessins de Albert...;](#)
4. [MÃdecine du cyclisme; par H. Judet,... et G. Porte,...; \[avec la collaboration de P. Miserez\];](#)
5. [La Consultation du mÃdecin; 30 nouvelles consultations; par le D J.P. de Mondenard;](#)
6. [La SantÃ des cyclistes; du loisir Ã la compÃtition; Dr J.-P. de Mondenard...;](#)
7. [Le Cyclisme, de l'Ãcole Ã la compÃtition; route et piste, Ã l'usage des jeunes cyclistes, coureurs, dirigeants, entraÃneurs, cadres techniques, prÃparation aux diplÃmes fÃdÃraux et bevet d'Ãtat; Daniel ClÃment...;](#)

8. Les Femmes à bicyclette à la Belle époque; présenté par Claude Pasteur;

9. Les Petites reines du Tour de France; Rémy Pigois;

10. Pratique du cyclisme; Peter Konopka;

Show more records

Swiss National Library: One hit.

1. Schwyzerland; [Kartenmaterial] :: Rigi - Vierwaldstättersee : Wanderkarte;

Die Deutsche Bibliothek: 62 hits.

1. Bike-Events ... : die Höhepunkte der Saison ; Bike plus

2. Trainingsplaner

3. Franken : [40 Spitzen-Touren, präzise Wegbeschreibungen, übersichtliche Tourendaten, Streckenskizzen + Höhenprofile] / Frank Klose

4. Fahrtechnik special : 10-Stufen-Plan zum Erfolg ; Super-Tips für Einsteiger ; die Geheimtricks der Pros / [Chefred. Olaf Beck (verantw.). Red./Autor: Michael Badichler]

5. Richtig Mountainbiken : Fahrtechnik und Training für Fitness- und Ausdauersportler / Urs Gerig/Thomas Frischknecht

6. Bike-Erlebnis Salzkammergut, Dachstein-Tauern-Region : [60 Mountainbike-Touren] / Wolfgang Heitzmann

7. Mountain bike training : for beginners and professionals / Achim Schmidt. [Transl. by Paul D. Chilvers-Grierson]

8. Schwarzwald : [50 Spitzen-Touren, präzise Wegbeschreibungen, übersichtliche Tourendaten, Streckenskizzen + Höhenprofile] / Frank Klose

9. Das @Mountainbike : das richtige Modell, das richtige Zubehör, die richtige Fahrtechnik / Brant Richards ; Steve Worland. [Aus dem Engl. von Olaf Sylla]

10. Weserbergland, Teutoburger Wald : [40 Spitzentouren, präzise Wegbeschreibungen, übersichtliche Tourendaten, Streckenskizzen + Höhenprofile] / Frank Klose

Show more records

Projecto HILT

Esquemas usados pelos depositários de informação seleccionados

Comunidade depositária	Esquema de classificação	Esquema de cabeçalhos de assunto	Tesouros
Arquivos (7)	Próprio (2)	Próprio (2) LCSH ¹ (1)	UNESCO (5) BET
Bibliotecas (11)	CDD ² (5) LC ³ (4) Próprio (1)	LCSH (8) MeSH ⁴ (1)	Próprio (2) UNESCO (1)
Museus (5)	CDU ⁵ (2) CDD (1) SHIC ⁶ (1)	LCSH (2) MeSH (1)	AAT ⁷ (2) MDA: <i>British Museums Materials Thesaurus</i> (2) Próprio (2) ROOT Thesaurus (1)
Colecções especiais (2)		LCSH (2)	UNESCO (1)
Serviços de informação em linha (15)	CDD (5) Próprio (4) LC (1) CDU (1)	LCSH (6) Próprio (4)	AAT (2) HASSET ⁸ (2) Próprio (2) MDA ⁹ (1) UNESCO (1)
Agências e iniciativas governamentais (2)	Guia de categorização de assuntos para a ciência e tecnologia da Defesa (1) Próprio (1)		UNESCO (1) Próprio (1)

(Nicholson e Wake, 2001, p. 65)

¹ *Library of Congress Subject Headings.*

² *Classificação Decimal de Dewey.*

³ *Library of Congress.*

⁴ *Medical Subject Headings.*

⁵ *Classificação Decimal Universal.*

⁶ *Social History and Industrial Classification* (um sistema de classificação largamente utilizado nos museus).

⁷ *Art and Architecture Thesaurus.*

⁸ *Humanities and Social Science Electronic Thesaurus.*

⁹ *Museums Documentation Association (British Museums Materials Thesaurus).*

Projecto HILT

Pontos fracos e fortes das linguagens analisadas*

Getty Art and Architecture Thesaurus (AAT)	
Pontos fortes	Pontos fracos
<ul style="list-style-type: none"> • Conforme com as normas internacionais para tesouros • Possui recursos e suporte • Actualização permanente • Mapeia-se com outros esquemas relevantes, como os LCSH • Desenvolvido para a comunidade da arte, incluindo bibliotecas, arquivos e museus • Excelente cobertura de arte, história da arte, arquitectura e artesanato • Cobre também todas as áreas da cultura material • Distingue os objectos descritos do seu conteúdo iconográfico • A sua versão em linha é fácil de pesquisar e percorrer 	<ul style="list-style-type: none"> • Fraca na cobertura temática de algumas áreas • Não inclui fenómenos naturais • Lenta na inclusão de novos conceitos em arte • Problemas com a terminologia local
Classificação Decimal Dewey (CDD)	
Pontos fortes	Pontos fracos
<ul style="list-style-type: none"> • Amplamente utilizada a nível internacional • Muito conhecida, de alto nível e familiar do ponto de vista do seu uso • Bons recursos • Hierarquia numérica consistente • Disponível noutras línguas • Disponível em formatos electrónicos 	<ul style="list-style-type: none"> • Não é actualizada com rapidez suficiente • Dificuldade em desenvolver áreas temáticas • Influência americana

*In: *HILT workshop breakout sessions: discussion issues* (2001). Glasgow, June 2001 [Em linha]. [Consult. 21 Maio 2003]. Disponível na WWW: <URL: <http://hilt.cdlr.strath.ac.uk/Dissemination/Workshop%20documents/BrkoutfullFINALword.doc>>

ANEXO 14 (cont.)

Library of Congress Subject Headings (LCSH)	
Pontos fortes	Pontos fracos
<ul style="list-style-type: none"> • Amplamente utilizada a nível internacional • Bons recursos • Abrangente, embora criada na base de exigências literárias • Mapeada com a CDD, o MeSH, o AAT (entre outros) • Disponível em formatos electrónicos 	<ul style="list-style-type: none"> • Criada numa base <i>ad hoc</i> para descrever documentos individuais mais do que domínios específicos do conhecimento • Estruturalmente imperfeita; não tem uma hierarquia consistente • Falta de especificidade em certos tópicos • Não é utilizada pela maior parte das pessoas de uma forma normalizada • Demasiado vasta e complexa para os técnicos que não sejam especialistas em indexação • Fraca acomodação para novos conceitos • Influência americana
Classificação Decimal Universal (CDU)	
Pontos fortes	Pontos fracos
<ul style="list-style-type: none"> • Ampla cobertura de assuntos • Dá resposta às necessidades de actualização • Disponível noutras línguas • Amplamente utilizada nas bibliotecas universitárias europeias • Prioridade à terminologia do Reino Unido / europeia • As notações podem representar assuntos com elevado grau de complexidade 	<ul style="list-style-type: none"> • Falta-lhe suporte em termos de ferramentas e recursos (comparativamente com os esquemas dos Estados Unidos) • Não está disponível em formatos electrónicos convenientes, apesar da sua versão em linha • A notação complexa pode ser difícil para os utilizadores
Tesouro da UNESCO	
Pontos fortes	Pontos fracos
<ul style="list-style-type: none"> • Conforme com as normas internacionais para tesouros • Razoáveis recursos e suporte • Disponível em várias línguas europeias • Pesquisável na <i>Web</i> • Simplicidade de utilização para indexar descrições de elevado nível • Aplicabilidade a vários assuntos • Análise detalhada de relações entre termos 	<ul style="list-style-type: none"> • Demasiado genérica, não suficientemente específica • Desenvolvimento variável das áreas temáticas • Não aceita sugestões de modificação ou adição

Projecto HILT

Conjunto de opções colocadas aos participantes (depositários de informação) no workshop de Junho de 2001*

I. Não fazer nada:

1. A inteligência artificial resolverá o assunto a seu tempo
2. Uma grande empresa – a Microsoft ou outra similar – resolverá o problema
3. Não é assim tão importante
4. Não há necessidade de encontrar uma solução
5. O problema é insolúvel

II. Estabelecer um processo humano com a intenção de chegar, em tempo, a uma solução: estabelecimento de um Agência de Terminologias, talvez baseada nas bibliotecas nacionais/mda/NCA

1. Estabelecer uma “task force” interdomínio e intersectorial para conduzir as comunidades a uma solução
2. Estabelecer uma Agência de Terminologias e uma “task force”

III. Adoptar uma forma de abordagem básica, gradual, com vista a futuros desenvolvimentos:

1. Criar, de modo gradual, caminhos cruzados (“cross walks”) de terminologia inter-serviços e inter-comunidades, construindo eventualmente uma solução parcial mas adequada
2. Tentar resolver o problema apenas para serviços electrónicos, talvez através da visão da web semântica
3. Providenciar facilidades de recuperação mais flexíveis para os utilizadores
4. Uma ou mais destas opções em conjunto (especificar, s.f.f.)

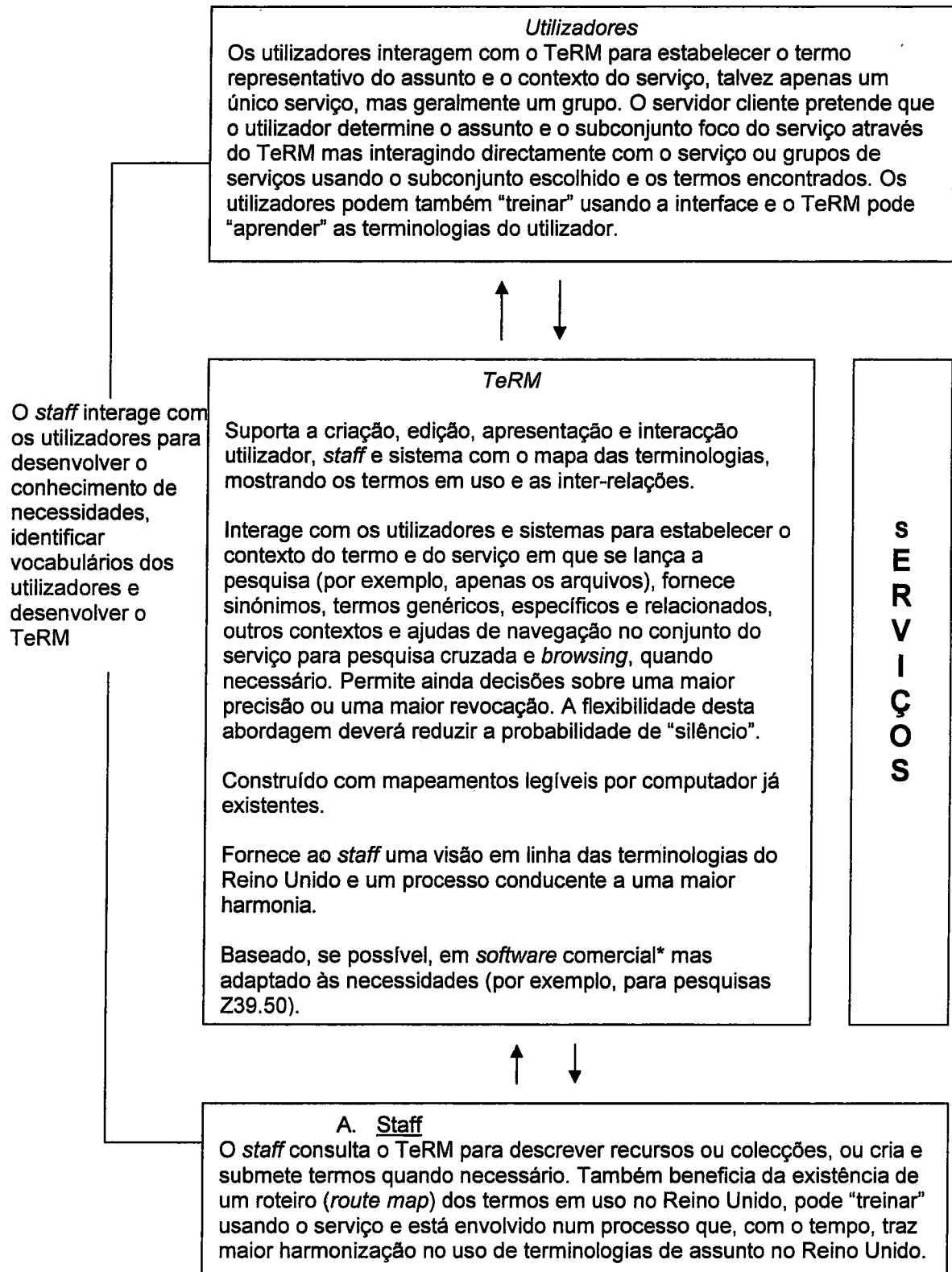
IV. Adoptar um único esquema:

1. Adoptar: LCSH / UNESCO / CDD / CDU / AAT / Outro esquema (diga qual) / Um novo esquema acrescentado ao já existente num dado serviço (especifique o esquema, s.f.f.)
2. Adoptar: LCSH / UNESCO / CDD / CDU / AAT / Outro esquema (diga qual) / Um novo esquema em vez do já existente num dado serviço (especifique o esquema, s.f.f.)
3. Adoptar um único esquema sem conversão retrospectiva de metadados / com conversão retrospectiva financiada pela organização / com conversão retrospectiva financiada a nível central

V. Alternativas de serviços de mapeamento:

1. Estabelecer um serviço de mapeamento, se possível com participação e suporte internacional, e gradualmente construir um mapeamento completo dos LCSH, UNESCO, CDU e AAT numa estrutura CDD. Incluir adaptações e extensões locais dos serviços mais importantes, como as bibliotecas nacionais. Usar o serviço internacional com os mapeamentos das adaptações e extensões do Reino Unido como modelo para outros países. Determinar e implementar o melhor modelo de financiamento e manutenção internacional.
2. Estabelecer um serviço piloto com a duração de 2 anos para medir os custos e benefícios tanto de um serviço completo como de todas as respostas alternativas detalhadas nesta página.

* Traduzido de Nicholson e Wake (2003), p. 66-67.

Projecto HILT - diagrama TeRM (*Interactive Terminologies Route Map*)*

*Nota: Podem ver-se exemplos em www.wordmap.com, com www.oingo.com e vivisimo.com.

* In NICHOLSON, Dennis (2002). [Tradução nossa]

Siglas e acrónimos utilizados

- AAT – Art & Architecture Thesaurus
- ANSI – American National Standards Institute
- AHRB - Arts and Humanities Research Board
- ASCII - American Standard for Codification and Interchange of Information
- BL - British Library
- BND – Biblioteca Nacional Digital
- BSO - Broad System of Ordering
- BSZ - Bibliotheksservice-Zentrum
- CAIRNS - Cooperative Information Retrieval Network for Scotland
- CARMEN - Context Analysis, Retrieval and Metadata : Effective Networking
- CC – Classificação Colon
- CDD – Classificação Decimal de Dewey
- CDU – Classificação Decimal Universal
- CEN – Comité Européen de Normalisation = European Committee for Standardization
- CENL - Conference of European National Librarians
- CERES - Californian Environmental Resources Evaluation System
- CERN - European Organisation for Nuclear Research
- CIDOC - International Committee for Documentation
- CIMI – Consortium for the Computer Interchange of Museum Information
- CLD - Collection Level Description
- CLEF – Cross-Language Evaluating Forum
- CLIP – Compatibilização de Linguagens de Indexação em Português
- CQL - Common Query Language
- CRM - Conceptual Reference Model
- DCMI - Dublin Core Metadata Initiative
- DDB – Die Deutsche Bibliothek
- DESIRE - Development of a European Service for Information on Research and Education
- DNER - Distributed National Electronic Resource

EELS - Engineering E-Library, Sweden
ELAG - European Library Automation Group
FAKTS - Facet Analytical Theory in Managing Knowledge Structure
FID - International Federation for Information and Documentation
FRBR - Functional Requirements for Bibliographic Records
FTP - File Transfer Protocol
GARE – Guidelines for Authority and Reference Entries
GARR - Guidelines for Authority Records and References
GDL – Glasgow Digital Library
GLIS - Geac Library Information System
HILT - High Level Thesaurus
HIP - Horizon Information Portal
HTML - Hypertext Markup Language
HTTP - Hypertext Transfer Protocol
ICOM – International Council of Museums
IFLA - International Federation of Library Associations and Institutions
ILL - Inter Library Loan
IP - Internet Protocol
IPM – Instituto Português de Museus
IPPAR - Instituto Português do Património Arquitectónico
ISADN – International Standard Authority Data Number
ISO – International Standard Organisation
JISC - Joint Information Systems Committee
LAC – Libraries and Archives of Canada
LCC – Library of Congress Classification
LCSH – Library of Congress Subject Headings
LEAF - Linking and Exploring Authority Files
MACS - Multilingual Access to Subjects
MARC - Machine Readable Cataloguing
MIME - Internet Media Type

MDA – Museum Documentation Association
MRF – Master Reference File
NGfL - National Grid of Learning
NLS – National Library of Scotland
NOF - New Opportunities Fund
OAI - Open Archive Initiative
OCLC - Online Computer Library Center, Inc.
OSI - Open Systems Interconnexion
OWL – Web Ontology Language
PICS - Platform for Internet Content Selection
PRECIS - Preserved Context Indexing System
PSP - Portal Suíço dos Periódicos
RAMEAU - Répertoire d’Autorité-Matière Encyclopédique et Alphabétique
Unifié)
RDF – Resource Description Framework
RDN – Resource Discovery Network
RERO – RéseauROmand
RSLP - Research Support Libraries Programme
RVM – Répertoire de Vedettes-Matières
Sbt - Sistema Bibliotecario Ticinese
SCONE – Scottish Collections Network
SHIC - Social History and Industrial Classification
SLAINTE - Scottish Libraries Across the Internet
SNL – Swiss National Library
SOA – Service Oriented Architectures
SOSIG – Social Science Information Gateway
SQL - Structured Query Language
SR - Search and Retrieval
SRW – Search/Retrieve Web Service
SUfi - Sottish University for Industry

SWB - Südwestdeutscher Bibliothekverbund
SWD / RSWK - Schlagwortnormdatei / Regeln für den Schlagwortkatalog
SYNTOL - SYNTagmatic Organization Language
TCP - Transmission Control Program
UBCIM - Universal Bibliographic Control and International MARC Core Activity
UKOLN - UK Office for Library and Information Networking
UNIMARC – Universal MARC format
UNISIST – Système Mondial d'Information Scientifique
URI - Uniform Resource Identifier
URL - Uniform Resource Locator
VIAF - Virtual International Authority File
W3C - WWW Consortium
WAIS - Wide Area Information Servers
WWW - World Wide Web
XML – eXtensible Markup Language
ZDB - Catálogo Colectivo Alemão de Publicações em Série
ZIG - Z39.50 Implementation Group