

# Tecnologias, Bibliotecas e Arquitecturas de Informação: dos Sistemas aos Objectos

Maria Inês Cordeiro

Biblioteca de Arte, Fundação Calouste Gulbenkian

Av de Berna, 45 A – 1096-001 Lisboa

Tel: 214578498

E-mail: ilopes@gulbenkian.pt

## RESUMO

A relação das bibliotecas com as tecnologias de informação atravessa uma importante fase de mudança, pelas transformações do universo da informação e do panorama tecnológico, duas faces da realidade incontornável que é a WWW. É abordada, em traços gerais, a situação actual das tecnologias em bibliotecas, face às necessidades de acesso, gestão e exploração de recursos de informação electrónicos, heterogéneos e distribuídos. A importância da interoperabilidade é sublinhada, como conceito e como aspecto técnico e estratégico, transversal à diversidade das questões relacionadas com a integração de recursos e serviços em rede. A função, significado e potencial de *standards* como o Z39.50, XML, metadados e RDF são realçados como sinais das alterações de paradigma que se vislumbram tanto para os sistemas de gestão de informação como para a realidade conceptual e técnica da WWW.

## PALAVRAS-CHAVE:

Tecnologias de informação e bibliotecas; sistemas integrados de gestão de biblioteca; interoperabilidade; Z39.50; XML; metadados; RDF.

## INTRODUÇÃO

Face às transformações operadas com a expansão da Internet, a última meia dúzia de anos é ainda um período relativamente curto no tempo das organizações para reflectir uma percepção clara dessas mesmas transformações e das suas implicações em aspectos estruturais dos modelos que enformam, ou virão a enformar, o conceito de biblioteca enquanto conjunto de recursos e de serviços. A mesma meia dúzia de anos não deixa de ser, por outro lado, um período já demasiado longo em termos da evolução do mercado de informação, na sua própria natureza e modo de funcionamento, com consequências imediatas sobre a articulação dos serviços de biblioteca nesse mesmo mercado.

Essas consequências sentem-se na descontinuidade de algumas funções, como a de controlo bibliográfico face ao universo de informação em rede [i]; e em desfasamentos a nível de determinados tipos de competências, muito visíveis no foro tecnológico, como as da gestão e preservação de recursos digitais. Pela natureza dos problemas, estes não são aspectos a que se possa fazer face com simples reajustamentos das estruturas existentes.

Essas descontinuidades e desfasamentos fazem parte do

processo de transformação e acabam por ser os seus impulsos que, como tal, são sobretudo emergências incompletas de questões mais globais e profundas. Sinal disso é a recente pulverização do conceito de “biblioteca” em novas representações sem que nenhuma delas consiga ainda transmitir senão facetas parciais, por vezes ambíguas, do que se pretende com elas comunicar. De um modo geral, novos termos vêm sucedendo ao simples termo “biblioteca”, até recentemente usado sem grandes ambiguidades para designar uma realidade estável – embora não necessariamente homogénea – com um modelo base suficientemente definido e claro.

Durante muito tempo as qualificações do termo surgiam para representar variações ou especificidades sobre esse modelo base, ditadas principalmente pelos tipos de organizações/missões em função das comunidades a que se destinavam (nacionais, sectoriais, universitárias, escolares, de empresa, ou especializadas por áreas temáticas...). Mais recentemente, todas as novas qualificações não surgem associadas a organizações/missões/comunidades, mas surgem principalmente para tentar significar duas coisas: por um lado, a necessidade de marcar a diferença relativamente aos paradigmas clássicos; por outro, uma associação explícita com as tecnologias de informação e comunicação, enfatizando aspectos baseadas na tipologia técnica dos recursos abrangidos ou na natureza das soluções para a realização técnica dos serviços.

É o caso dos termos biblioteca “electrónica”, “virtual”, “digital”, aliados a outros designativos mais metafóricos como “sem paredes”, “sem fronteiras”, etc, ou ainda, mais recentemente, a biblioteca “híbrida”...sugerindo conciliação ou, finalmente, biblioteca *gateway* para enfatizar a importância do acesso “acreditado” [ii]. Quando estes termos são usados estaremos todos a falar do mesmo? Na maior parte das vezes não e o fenómeno é esse mesmo: a linguagem é experimental e exploratória enquanto os conceitos são difusos [iii].

A razão de introduzir aqui este aspecto da terminologia – em si própria pouco importante – é apenas por lhe achar alguma relevância ilustrativa da nossa (in)capacidade actual para preencher o sentido e dominar o significado das transformações. Que elas são mais extensas e profundas que as de outras evoluções anteriores, é um dado que o actual mercado da informação não ilude, na complexidade de todos os seus aspectos não só tecnológicos mas também sociais, económicos e legais. Que essas transformações são susceptíveis de envolver a alteração de paradigmas e dificultar identidades e

modelos, é um aspecto que a diversidade e ambiguidade semânticas, que actualmente rodeiam o conceito de biblioteca, revelam.[iv]

Essa diversidade e ambiguidade permanecerão até uma reconciliação mais completa com o simples termo/conceito de “biblioteca”, que significará antes de mais libertar a sua filosofia geral da circunstância ou dos modos particulares da sua realização, face aos mundos analógico e digital. Por um lado esses modos não se excluem nem antagonizam, antes se evidenciam mutuamente. Por outro, a relação com o mundo digital não é hoje, em 2001, uma opção. É uma evolução natural e “ecológica” da própria existência de qualquer biblioteca, como acontece em muitos outros tipos de organizações e áreas de actividade.

Libertar a “biblioteca” de qualificativos que a “distingam” por associação às tecnologias de informação é libertar o próprio conceito de biblioteca e possibilitar uma visão mais abrangente e profunda das questões que se colocam às bibliotecas em geral, no médio e longo prazos, independentemente das circunstâncias, particularidades ou dos projectos individuais de cada uma.

Nessa visão abrangente é pertinente, antes de mais, um olhar sobre a relação das bibliotecas com a tecnologia e perceber o que está hoje a mudar nessa relação, já que a mesma não é nova. Em seguida, é importante destacar alguns dos que são hoje os movimentos de fundo do panorama tecnológico, com capacidade para vir a influir de forma decisiva e duradoura na natureza e modo de funcionamento do mercado de informação, de que as bibliotecas são parte.

## TECNOLOGIAS E BIBLIOTECAS

Ao longo de três décadas que as bibliotecas vêm utilizando as tecnologias de informação para controlar e gerir as suas colecções e serviços. Os sistemas integrados de gestão de biblioteca criaram, sobretudo desde os anos 80, o seu espaço de mercado, evoluindo com as exigências desse mercado, cada vez mais especializado em termos de requisitos normativos de âmbito nacional e internacional. Duma forma geral, são hoje sistemas bastante sofisticados e maduros, que respondem a necessidades de uma comunidade muito alargada mas com uma sólida base comum quer em matéria de arquitecturas de informação quer no que respeita a modelos funcionais.

A estabilidade dessas arquitecturas e modelos correspondeu durante muito tempo à própria estabilidade do mundo da informação impressa com o qual o modelo base da biblioteca consolidou a sua articulação. Nesse contexto, os sistemas de gestão de biblioteca realizam a *informatização* de rotinas, processos e estruturas de informação que, até hoje, seguem basicamente o mesmo modelo que já existia nos sistemas manuais. Por essa razão, e até muito recentemente, a “informatização” de bibliotecas foi sobretudo uma aplicação de produtos acabados de tecnologias, maioritariamente soluções *standard* “chave na mão”, cuja absorção nas instituições se foi efectuando sobretudo na “óptica do utilizador”.

Do ponto de vista tecnológico, estes sistemas evoluíram com o desenvolvimento da tecnologia dos sistemas de bases de dados, passando dos sistemas de ficheiros para os modelos relacionais, nos anos 80, e introduzindo aspectos relativos à interconexão, com a vulgarização das redes de dados, nos anos 90. [v] De certa forma, a relação da biblioteca com a tecnologia acompanhou essa evolução mas sempre principalmente na óptica do utilizador, não alterando substancialmente nem os seus modelos nem os níveis de competência e actividade nessas áreas. As tecnologias para gestão ou utilização da informação continuaram a ser relativamente exteriores e complementares face ao que tem sido considerado as competências fundamentais e específicas das bibliotecas em matéria de organização da informação.

Actualmente, do mesmo modo que se sentem descontinuidades e defasamentos na função das bibliotecas face ao mundo da informação digital em rede, também os sistemas de gestão de biblioteca já estão longe de encerrar soluções completas para todos os desafios que se lhes colocam. Por outro lado, em geral esses mesmos sistemas ainda têm quase todos uma arquitectura técnica que data dos anos 80 e só pontualmente introduzem tecnologia mais recente da geração Internet [vi]. Não é expectável que a indústria faça, rapidamente, investimentos vultuosos e radicais em novos produtos tecnologicamente mais actualizados para bibliotecas, quando a própria tecnologia não é estável, quando estão em aberto diversos aspectos do que serão os modelos funcionais futuros e quando a área de mercado a que esses produtos se dirigem não é propriamente a mais atraente do ponto de vista comercial. [vii]

Mas também não se trata agora simplesmente de actualizar a tecnologia dos sistemas integrados de gestão de biblioteca, ou de acrescentar novas funcionalidades aos modelos existentes. O lugar e o fundamento das tecnologias nas bibliotecas já ultrapassa em muito a questão da “informatização” de processos cuja concepção é, aliás, muito anterior à tecnologia.

As grandes alterações tecnológicas a que hoje assistimos são determinantes para quase todas as áreas de actividade, principalmente porque modificam os meios e modos que influem na performance dessas actividades. No caso das bibliotecas são mais do que isso, pois são alterações que afectam a natureza intrínseca do próprio objecto da sua actividade: a informação, as suas estruturas, os seus suportes, os seus modos de produção, aquisição, transmissão, arquivo e utilização.

Basta pensarmos, relativamente aos requisitos de domínio da tecnologia para efeitos de gestão dos objectos de informação, como são enormes e estruturais as diferenças entre os objectos analógicos e os digitais, e como se tornam mais complexas quando a existência dos objectos digitais se realiza principalmente em rede.

Não só porque a natureza do objecto se diversifica e modifica substancialmente, mas também porque na

dinâmica actual da informação e da tecnologia Web as alterações se sucedem e estão longe de oferecer um quadro estabilizado, torna-se crucial uma relação mais estreita – e mais “natural” – das bibliotecas com as tecnologias de informação. É a convergência das tecnologias com a gestão de informação, que se apresenta agora com maior premência. [viii] Por várias razões.

Primeiro, porque é improvável que, pelo menos num futuro próximo, existam soluções comerciais equivalentes, em conceito e abrangência, ao que são hoje os actuais sistemas integrados de gestão de biblioteca, para responder a todas as necessidades, já que a tendência é para uma multiplicidade de soluções, entre si complementares, voltadas não tanto para modelos monolíticos e autosuficientes, mas sobretudo para a federação de diversos recursos e serviços em rede.

Em segundo lugar, porque tendo as bibliotecas que fazer face às questões mais prementes na modernização dos serviços ao público, a estratégia adequada é a de ir actuando de forma ágil e pragmática na integração de soluções complementares, sem esperar que existam “off the shelf” instrumentos completos.

Por último, a própria tecnologia está a introduzir modificações fundamentais nas arquitecturas dos objectos de informação, com repercussões a nível dos modelos de serviços de intermediação, de pesquisa e recuperação ou de manutenção de recursos.

Todas estas razões tendem a que a biblioteca tenha que diversificar competências e assumir progressivamente um nível de actividades tecnológicas que até há algum tempo lhe era relativamente exterior, ou lateral [ix]. O que não é fácil, imediato, ou isento de equívocos.

A relação mais directa com novos aspectos da realidade tecnológica iniciou-se há cerca de dez anos, abrindo a época dos chamados “projectos-piloto”, apoiados por políticas de incentivo à inovação de serviços de biblioteca [x]; de há cinco a esta parte vivemos também na época das chamadas “iniciativas” – este um termo interessante, a expressar o não alinhamento dessas actividades com a filosofia e a natureza das realizações institucionais clássicas.

Embora sem dúvida impulsionadores da inovação, em boa parte dos casos esses projectos e iniciativas têm-se desenvolvido nas margens da vida corrente das bibliotecas, não raro constituindo primeiro oportunidades para as áreas das tecnologias, para só depois chegarem às áreas profissionais da gestão da informação: esta colocação das questões em dois tempos e dois planos não tem facilitado a comunicação, essencial para uma convergência cada vez mais necessária a ambas as partes.

Mesmo contando com a participação das bibliotecas, ou sendo promovidos por elas, esses projectos são-no ainda como um “adicional” à sua realidade existente, como crescimentos laterais, muitas vezes efémeros, que ainda pouco influem na sua cultura profissional [xi] [xii].

Uma das principais dificuldades em se ultrapassar esta situação advém do facto de as bibliotecas terem necessariamente – e simultaneamente – que canalizar prioridades, recursos e atenções para o curto prazo, onde não se pode roubar o lugar das actividades e serviços tradicionais, cuja necessidade se mantém. Outra das dificuldades é inerente ao ciclo da própria geração dos conteúdos profissionais: sendo as ciências da informação/biblioteconomia uma área com fraca identidade académica e com uma expressão diluída por várias áreas em termos de campos de investigação, caberia ao mercado impulsionar a renovação dos conteúdos profissionais, colocando novas exigências na procura.

Mas o mercado são as próprias instituições, na sua maioria organizações não lucrativas, cuja evolução organizacional e tecnológica é condicionada por múltiplos factores e contingências que não são da mesma natureza dos que determinam, nas empresas, as estratégias de sobrevivência pela competitividade e rentabilização de recursos. Apesar de crucial, o envolvimento profissional mais profundo das bibliotecas com as tecnologias acaba, assim, por ser usualmente lento, mesmo quando existe a possibilidade de fazer uso de algumas tecnologias de ponta.

Se é certo que não é fácil acelerar a mudança de cultura profissional na prática, é pelo menos desejável que exista um acompanhamento atento ao que se passa no ambiente tecnológico, numa atitude de procurar identificar, compreender e intervir naquilo que é mais significativo e que poderá ter efeitos estruturantes no futuro. Cabe aqui antes de mais realçar que nos encontramos numa época dos desenvolvimentos normativos *de facto* [xiii]. Ou seja, o aparecimento e implantação de modelos e *standards* tecnológicos e de gestão de informação da época da Web não segue a mesma filosofia, trâmites e canais habituais, ou a que as instituições estejam habituadas. [xiv] As iniciativas que conduzem a novos *standards* surgem a partir de instâncias novas ou não oficiais num ambiente normalmente aberto a qualquer tipo de entidade, o que não acontecia anteriormente. Por outro lado, essa abertura, que é própria da natureza da Internet, acarreta consigo uma rapidez de processos que é também inusitada.

Estes aspectos contrastam com a tradicional lentidão e ortodoxia dum sector como o nosso, onde estamos habituados a esperar para cada detalhe uma norma, para cada norma sua instância e de cada instância um tempo institucional e administrativo que, não raro, interpõe lapsos com que as necessidades práticas não se compadecem. Modificar esta situação é introduzir uma atitude proactiva, mais flexível e indagadora, em planos mais alargados [xv].

É neste sentido que se torna importante observar e conhecer algumas vertentes e tendências da evolução das tecnologias de informação que, com a expansão da WWW, são susceptíveis de ter fortes implicações no paradigma dos sistemas de informação em geral [xvi] e nos das bibliotecas, em particular, nomeadamente nos aspectos que podem ter a ver com a descoberta e recuperação de recursos de informação em rede, assim como com as suas qualidades funcionais, necessárias para a sua gestão.

Na turbulência geral, algumas dessas tendências envolvem inovações que se podem confundir com simples novidades, quer pelo aparente modismo que as rodeia quer pelas promessas entusiásticas, por vezes demasiado simplistas, que à sua volta se geram. Tal pode ser a situação de normativos como o XML e os modelos de metadados, surgidos para solucionar muitas das novas questões que se colocam à gestão e exploração de recursos electrónicos heterogéneos e distribuídos.

O problema base que tais desenvolvimentos visam é um problema de *integração*. À urgência na procura de soluções para integração de sistemas, de objectos, e mesmo de tecnologias, no ambiente da Web, corresponde a importância de um conceito chave - o conceito de interoperabilidade - que é vital porque é subjacente e transversal à diversidade das questões emergentes da proliferação de informação em rede.

### **A INTEROPERABILIDADE COMO CONCEITO CHAVE**

Genericamente, interoperabilidade designa as qualidades de um sistema, produto ou objecto, que lhe permitem interagir com outros de forma transparente. Significa isto a possibilidade de reduzir, ou eliminar, o esforço adicional que a diversidade acarreta, através de funcionalidades de exploração que disponibilizem sistemas, produtos ou objectos diversos como se de um só se tratasse. É fácil perceber que o conceito se liga a objectivos de eficiência e eficácia, através da optimização de processos que utilizam recursos independentes. No âmbito dos sistemas de informação essas qualidades e objectivos traduzem-se num aspecto essencial que é a capacidade de *integração*.

Não sendo novo, nem em termos puramente tecnológicos nem em termos mais globais da filosofia dos sistemas de informação, o conceito assume hoje uma dimensão e repercussões sem precedentes. Até muito recentemente interoperabilidade e integração eram preocupações apenas ao nível dos sistemas de gestão de informação, para comunidades de sistemas com natureza idêntica e normalmente com as mesmas finalidades principais.

Com o crescimento exponencial da WWW novos problemas se têm vindo a colocar relativamente à acessibilidade, utilização e gestão dos recursos de informação, a uma escala e com uma heterogeneidade que exigem soluções mais alargadas. Não só cresceu o imperativo da interoperabilidade entre sistemas como se evidencia cada vez mais a necessidade de haver qualidades que viabilizem a interoperação ao nível dos próprios objectos individuais de informação – recursos de qualquer natureza e tipologia que povoam a WWW, através de páginas web ou a elas ligados. [xvii]

A expansão da Web e o seu sucesso tornaram cada vez mais visíveis as suas próprias limitações: grande assimetria de conteúdos com enorme irregularidade de estruturas, e formas de organização rudimentares que tornam frágeis e dificilmente sustentáveis as soluções de integração. Neste contexto surgem normativos, como o XML e modelos de metadados, abrindo um espaço novo entre a tradicional cultura dos sistemas de bases de dados e

a cultura tecnológica do seu mundo exterior e posterior, o da Web.

Embora as questões de descoberta, pesquisa e recuperação de informação na Web sejam as mais imediatamente notadas e as mais frequentemente abordadas, o âmbito destes novos normativos não é apenas o espaço das questões relativas à tecnologia dos motores de pesquisa ou dos sistemas de gestão de informação; é sobretudo o espaço em que se debatem questões genéricas de arquitecturas de informação que pretendem ser mais universais e independentes dos agentes que as produzem e utilizam. Com esta mudança de perspectiva, esse espaço constitui um plano totalmente novo na relação entre a tecnologia, a gestão de informação e todos os factores e diferentes tipos de agentes que intervêm na produção, transmissão e uso de informação. Esses factores e agentes não são só tecnológicos.

É neste quadro que as questões da interoperabilidade se tornam mais cruciais e que o conceito sai do círculo técnico relativamente restrito em que residia infiltrando-se por todas aquelas camadas da nossa realidade que de alguma forma têm a ver com informação digital em rede. A interoperabilidade não passou só a ser encarada como qualidade central das soluções em tecnologias de informação, logo nos produtores [xviii], como passou também a constituir matéria de interesse e preocupação em instâncias muito diversas e transversais, que são de ordem organizacional, económica e cultural [xix], inscrevendo-se expressamente nas estratégias de sectores de actividade e de organizações [xx], nas preocupações dos programas de política para o desenvolvimento da sociedade da informação, a nível regional, nacional ou internacional e ainda nas prioridades das instâncias normativas [xxi].

### **INTEROPERABILIDADE E SISTEMAS DE GESTÃO DE INFORMAÇÃO**

Do ponto de vista tecnológico o conceito começou por se vulgarizar a partir dos anos 80 com o desenvolvimento da filosofia dos sistemas abertos e a elaboração do modelo de referência OSI [xxii], surgido para orientar a indústria no sentido de reduzir os inconvenientes do carácter proprietário dos diversos componentes técnicos dos sistemas - plataforma de HW, sistemas operativos, gestores de bases de dados ou aplicações – que dificultam a sua interconexão. O modelo completo foi raramente implementado quer por ser complexo quer por ter tido um longo desenvolvimento teórico que foi em certa medida ultrapassado pelos avanços da tecnologia, nomeadamente a tecnologia Internet que assume um carácter de normalização *de facto*. No entanto, na sua essência, o modelo OSI permanece ainda como um modelo fundacional: é uma abordagem que visa reduzir a complexidade controlando as questões da diversidade por camadas ou níveis de interacção, numa forma estruturada que permite identificar e resolver, com o mínimo de dependências possíveis, as questões de cada nível.

A conotação quase estritamente informática do conceito de interoperabilidade é, assim, uma herança dos anos oitenta, intimamente ligada ao desenvolvimento de protocolos de rede. Entre estes merecem especial referência os que se situam ao nível das aplicações, que é onde as questões de

partilha de informação se colocam, e onde o Z39.50 (ISO 23950) [xxiii] se inclui. Sobre o objectivo e modo de funcionamento deste protocolo não será necessário entrar em detalhes, já que ele é, genericamente, conhecido no nosso âmbito.[xxiv] A sua menção aqui serve apenas para exemplificar – melhor do que outros que também poderiam ser mencionados, caso dos que se relacionam com o comércio electrónico – os tipos e níveis de exigências que os objectivos de interoperabilidade colocam às estruturas e conteúdos da informação.

A interoperabilidade entre sistemas requer não só *mecanismos que formalizam a comunicação entre sistemas* – os protocolos de rede – mas também qualidades comuns em matéria de sintaxes e semânticas dos dados a que possam ser aplicados esquemas de concordância comuns. Se na concepção de cada sistema ao nível das arquiteturas de informação e estruturas de dados nada houver de comum que permita estabelecer qualquer nível de concordância, então o protocolo, enquanto mecanismo, não é implementável. Por outro lado na medida em que essa concordância for facilitada pela adopção de normas comuns, nos vários níveis de representação que a informação tem nos sistemas – assim é maior a exequibilidade da implementação dum protocolo ao nível da aplicação e melhores são as perspectivas de uma interoperabilidade mais cabal, que pode ser explorada em funcionalidades mais complexas.

O que explica a quase exclusiva utilização do protocolo Z39.50 nos sistemas de informação bibliográfica não é senão a existência de uma larga comunidade de sistemas cujas arquiteturas de informação e estruturas de dados assentam numa sólida base normativa a vários níveis:

- o nível ontológico das unidades de representação (que tipos e níveis de entidades lógicas de informação),
- o nível conceptual das formas de representação (modelos de organização dos elementos de informação, por exemplo ISBD e códigos de catalogação);
- o 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);
- as normas de codificação de caracteres.

Se, para além desta “ossatura” a concordância entre sistemas for também possível ao nível do próprios conteúdos de elementos de informação (por exemplo a utilização dum mesmo ficheiro de autoridade de nomes ou instrumento de controlo das formas de acesso por assunto) então a interoperabilidade é mais conseguida, mas neste caso por factores de gestão de conteúdos a um nível que ultrapassa os formalismos da “legibilidade” por computador.

Os aspectos de menor sucesso do Z39.50 revelam principalmente dificuldades e problemas que não são, na sua maior parte, de ordem tecnológica. No caso das bases de dados bibliográficas, e apesar do seu elevado grau de normalização, salientam-se diferentes tipos de idiosincrasias capazes de afectar o nível de interoperabilidade: por um lado, diferenças de parametrização ao nível dos sistemas de gestão de informação.[xxv] A outro nível, o dos conteúdos, é vulgar

existirem diferenças nas representações para os mesmos objectos de informação: diferenças de granularidade e conteúdo das descrições bibliográficas, diferenças nos vocabulários de pesquisa (pontos de acesso) ou ainda diferenças na língua, neste caso para todos os elementos descritivos ou de acesso que surgem na língua do catálogo. [xxvi]

A interoperabilidade implica, assim, aspectos sintácticos e semânticos sendo que estes têm vários níveis. O Z39.50 é aqui referido como paradigmático dessa realidade: o protocolo em si resolve uma pequena parte, a sintáctica; os perfis de implementação resolvem a maior parte, através de convenções adoptadas por comunidades de sistemas para “limitar” os critérios de implementação do protocolo e resolver o quadro principal da interoperabilidade semântica: ou seja, o perfil estabelece um conjunto de especificações que limitam as opções que a norma oferece, em função dos requisitos técnicos e funcionais adequados a uma comunidade de sistemas.

Os problemas que não podem ser resolvidos pelo protocolo e pelo perfil implementado são problemas a montante, acima referidos como as idiosincrasias de cada sistema ou catálogo: estes só podem ser ultrapassados por acordos ou políticas comuns de normalização em matéria de parametrização de sistemas ou de gestão intelectual da informação.

Inicialmente, segundo o modelo em que originalmente foi concebido, o protocolo funcionou para a interrogação remota de servidores individuais, em que é possível, embora possa ser trabalhoso, configurar um cliente Z39.50 de forma a atender às particularidades do sistema de pesquisa de cada servidor. A partir do momento em que o protocolo é usado para pesquisar simultaneamente em vários servidores – num modelo também designado de *descoberta de recursos* ou *gateway* – as questões de semântica tornam-se mais cruciais, quer ao nível dos perfis de implementação (numa comunidade de servidores congéneres podem ser adoptados perfis ligeiramente diferentes) quer ao nível das questões a montante do protocolo. Por essa razão, as definições dos mais recentes perfis do Z39.50 – caso do *Bath Profile*<sup>xxvii</sup> – são suficientemente específicas para prever, para cada tipo de pesquisa/conjunto combinado de atributos do perfil, um comportamento conhecido e uniforme dos servidores que o implementam. [xxviii]

Serve esta explicação a propósito do Z39.50 para clarificar a verdadeira natureza e dimensão das questões que dificultam a interoperabilidade, e que não são nem principalmente informáticas, a nível do protocolo, nem só técnicas, a nível das estruturas de dados; são também organizacionais. A nível informático é por vezes referida a inadequação do protocolo para o ambiente tecnológico da Web [xxix]. Mas é também um facto que funciona, na WWW e fora dela, e não tem uma alternativa à vista, baseada em tecnologia puramente Web. [xxx]

Por essa razão, e porque é agora, mais do que nunca, que o ambiente de rede exige soluções que funcionem, o interesse sobre o Z39.50 redobrou nos últimos três anos, ao mesmo tempo que prosseguem os esforços de alinhamento com a tecnologia Web [xxxi], por exemplo a codificação das mensagens Z39.50 em XML, utilização com

metadados Dublin Core, etc. [xxxii]

O potencial de funcionalidade do Z39.50 tem os seus requisitos e os seus custos: é um protocolo complexo em termos de implementação, porque exige conhecimentos de arquiteturas de informação específicas para nichos relativamente circunscritos de mercado - como as bibliotecas, museus, informação governamental, etc., estas de facto as áreas onde existem, por natureza, bases de dados com informação estruturada segundo normas comuns e vocacionadas para disponibilização pública.

O facto de ser exigente nos requisitos para dar bons resultados, de ser de implementação complexa e de não atrair as atenções comerciais em nada prejudica a qualidade do serviço que presta e em nada justifica que não continue a ser, no nosso âmbito, explorado, enquanto não houver alternativas de serviço equivalente. Podemos mesmo considerá-lo um bom exemplo do que R. Kling refere como o “paradoxo do bibliotecário”. [xxxiii]

Tal como o modelo OSI continua a ser fundamental enquanto modelo genérico, não só para a interconexão mas inclusive para a modelação de dados visando a interoperabilidade [xxxiv], também o Z39.50 mantém o mérito de ter avançado com modelos para os processos de recuperação de informação. Para além dos modelos, dá-nos a experiência prática das questões de interoperabilidade semântica, as quais são essencialmente as mesmas, qualquer que seja o tipo de objecto – bases de dados, repositórios ou objectos de informação individuais - ou a tecnologia usada para as resolver. Este é um aspecto da maior importância, para a compreensão do que se passa actualmente no contexto da Web com a construção da chamada Web semântica.

### **A WORLD WIDE WEB: UM ESPAÇO DE DUAS CULTURAS TÉCNICAS**

Com o aparecimento da Internet as questões de conectividade entre sistemas ficaram largamente facilitadas: a *suite* de protocolos TCP/IP veio simplificar as transacções de comunicação de várias camadas do modelo OSI, ao nível do transporte, ao mesmo tempo que os standards HTTP/HTML revolucionaram a publicação de informação em rede, através de soluções universais, abertas e muito pouco exigentes.

Embora situados ao nível da aplicação, os *standards* HTTP/HTML não estão virados para as questões da interoperabilidade, mas sobretudo para os aspectos de acesso e apresentação de conteúdos. Para as bases de dados estes *standards* significaram mudanças visíveis e atractivas do ponto de vista do utilizador, mas limitadas a diferenças de apresentação: muitos sistemas passaram a ser providos de interfaces HTML, e como tal utilizáveis através de um *browser*. A visibilidade das bases de dados é também maior já que têm necessariamente que estar associadas a páginas web que as anunciam. Para além disto, que é apenas a superfície dos sistemas na Web, só existe interoperabilidade e integração funcional se, concomitantemente, forem usados outros mecanismos, como o Z39.50. Sem isso, os sistemas estão na Web mas apenas passíveis de uma utilização individual, sendo que o utilizador tem que conhecer e interpretar cada um no seu interface, nos seus critérios, nas suas funcionalidades de

pesquisa e nos modos de apresentação de resultados próprios.

Anteriormente à WWW a realidade dos recursos em rede consistia em sistemas de bases de dados que disponibilizam acesso remoto. Com o aparecimento da WWW não são apenas bases de dados que estão em rede mas também objectos individuais de informação ou repositórios de objectos. Passou a existir, por isso, a necessidade de aceder, recuperar e usar tanto a informação residente em bases de dados como em páginas Web, suas partes ou quaisquer tipos de recursos electrónicos a elas ligados. [ xxxv]

A Web tornou-se, assim, um espaço onde coexistem dois mundos diferentes, com duas culturas distintas: a cultura das bases de dados e a cultura da tecnologia Web. Não sendo antagónicas são, no entanto, conceptualmente e tecnicamente diferentes [xxxvi] e, até agora, representam universos de informação também diversos, mesmo quando “aparecem” combinadas como, por exemplo, numa base de dados com interface Web ou referenciando recursos da WWW. Mesmo combinados, esses dois mundos exibem claramente as suas diferenças:

- os recursos das bases de dados são contidos e altamente estruturados, a todos os níveis necessários: sintaxe e semântica dos dados, relações, separação entre registo e apresentação de dados, etc. existindo claramente a distinção entre níveis lógicos e níveis físicos nas estruturas de informação; a informação na Web, nomeadamente HTML, compreende conjuntos ilimitados de recursos com estruturas de dados muito incipientes, maioritariamente vocacionadas para a apresentação e hiperligações, praticamente não existindo um nível de estrutura lógica abstracta;
- as bases de dados possuem mecanismos que lhes garantem a integridade e consistência, assim como técnicas de armazenamento de dados e de garantia de salvaguarda; na Web não existem relações estruturais entre os recursos de informação, apenas hiperligações; não há garantia de integridade e permanência dos recursos;
- as bases de dados detêm serviços de pesquisa próprios eficientes e optimizados para os seus conteúdos, mas estes estão fora do alcance dos serviços de pesquisa Web; por seu turno, a acessibilidade dos recursos disponíveis directamente na Web depende de serviços que lhe são externos e alheios, variando entre os que garantem uma larga cobertura mas com qualidades de recuperação deficientes (serviços automatizados dos motores de busca) e os que garantem boas qualidades de recuperação mas limitados a uma cobertura reduzida (normalmente os serviços baseados em selecção e indexação humanas ou mistos).

Em termos de conteúdos ambos são igualmente importantes e não se substituem: as bases de dados contêm informação estruturada para comunidades e finalidades específicas e incluem repositórios relativos a grandes massas de recursos que não estão, e provavelmente não vão estar, disponíveis na Web. Por outro lado, o espaço Web contém vastos recursos de informação, sobretudo

corrente, que na sua maior parte não estão, nem vão estar disponíveis doutra forma, nomeadamente impressa.

A integração de bases de dados e objectos ou repositórios de objectos digitais requer a construção de serviços automatizados de mediação que alteram a tradicional configuração cliente/servidor em arquitecturas *multitier*: nestas acrescentam-se soluções intermédias (middleware) que têm a função de transformar as estruturas de dados existentes nos sistemas servidores nas estruturas desejadas a nível dos clientes, por exemplo HTML. Caso de trate de um serviço de mediação comum para acesso à informação de vários servidores, usualmente com estruturas de dados e funcionalidades diferentes, esse serviço terá que contemplar todas as traduções necessárias das *queries* e das respostas para as integrar “virtualmente” de acordo com um único modelo lógico de dados, que é o que é oferecido ao cliente, num interface de pesquisa único. [xxxvii]

Por outro lado, a integração de outros recursos individualmente e directamente disponíveis na Web, é dificultada pela falta de estruturas de representação comuns e de um mínimo de estrutura lógica que possa ser processada por computador por agentes independentes. As capacidades dos motores de pesquisa Web têm sido sucessivamente melhoradas com expedientes de vária natureza, de forma a poderem minorar as dificuldades de descoberta, identificação e descrição da diversidade de recursos existentes [xxxviii]; no entanto, todos eles se deparam com sérias limitações que a sua tecnologia, só por si, dificilmente resolve se os próprios recursos não fornecerem condições para uma mais sofisticada manipulação de dados por computador, nomeadamente em termos de estrutura lógica e valor semântico.

Existe assim a necessidade de dotar a Web de *standards* abertos capazes de melhorar a arquitectura [xxxix], a forma de representação dos dados e o conteúdo dos próprios objectos individuais de informação. [xl]

## **INTEROPERABILIDADE E OBJECTOS DE INFORMAÇÃO: XML, METADATA E RDF**

É neste contexto que se desenvolve o conceito de *Web semântica* [xli], constituindo uma visão em que os dados na WWW – documentos ou partes de documentos – estejam definidos, expressos e interligados de tal forma que possam ser usados por computadores não apenas para efeitos de apresentação mas também para processamento automatizado para outros fins, nomeadamente integração funcional e reutilização de dados entre diferentes aplicações. Neste conceito, a audiência da Web não são apenas pessoas mas também computadores.

O XML e os metadados constituem dois dos principais elementos dum largo espectro de especificações técnicas que têm surgido nos últimos tempos para responder a esta necessidade - que é a necessidade de arquitecturas de informação adequadas a uma evolução da Web, aberta mas sustentada e virada para a interoperabilidade - após o que foi designado por “ciclo do HTML”. [xlii]

## **XML – uma linguagem para estruturar conteúdos e melhorar o processamento de dados**

A linguagem XML (*Extensible Markup Language*) [xliii] surge basicamente para ultrapassar as limitações do HTML em termos de estruturação dos dados que constituem os documentos Web. Essas limitações decorrem de um número limitado e não expansível dos elementos que os compõem, e também do facto de se tratar de uma linguagem quase exclusivamente voltada para a definição das características de apresentação (*layout*) de documentos. O HTML é assim essencialmente uma linguagem para publicação com sérias limitações no que respeita à manipulação automática, nomeadamente para efeitos de recuperação de informação, já que não possibilita a identificação e descrição dos conteúdos que os compõem. [xliv]

Com a sua primeira versão publicada em 1998, a linguagem XML, é normalmente apresentada como uma versão simplificada da primeira linguagem da família *markup*, o SGML (*Standard Generalised Markup Language*), publicado em 1986, mas não tão simplificada como o HTML. Contrariamente a estas duas normas, o XML não se refere apenas a uma determinada norma, mas constitui a principal de um conjunto de especificações relacionadas entre si, que continuam a ser desenvolvidas, para prever toda a funcionalidade que se pretende dos documentos Web.

Tal como as anteriores linguagens *markup*, o XML é uma representação textual de dados definida por *elementos*, mas mais rica que o HTML já que prevê a representação de estruturas mais detalhadas e significativas, em que o documento se pode descrever a si próprio em termos do seu conteúdo. Como o próprio nome indica, o XML é extensível, o que significa que as etiquetas inseridas para indicar os elementos que compoem a estrutura do documento podem ser especificadas no número necessário a cada conteúdo, da mesma forma que cada elemento se pode desenvolver em sub-elementos e atributos, até um nível de especificidade que é arbitrário.

Opcionalmente, os documentos XML podem conter também, ou estar associados a uma descrição da sua “gramática” (etiquetas, sub-elementos, atributos, etc) definida numa DTD (*Document Type Definition*) ou num *XML Schema*. Estas definições permitem que a mesma “gramática” possa ser válida para vários documentos e acordada para as necessidades de comunidades específicas. Por outro lado, a existência dessas gramáticas viabiliza a validação da sintaxe dos documentos, assim como facilita o seu processamento automatizado.

Um outro aspecto fundamental é o facto de a definição da forma de apresentação dos documentos ser independente da estrutura do seu conteúdo: essa definição cabe a um ficheiro independente formulado de acordo com regras próprias para o efeito - CSS (*Cascading Style Sheet*) ou XSL (*Extensible Style Language*). Isto permite que ao mesmo documento possam ser aplicadas diferentes definições de apresentação, o que dá outra maleabilidade à manipulação de conteúdos.

Por fim, é de referir a possibilidade de “tradução” de dados de e para outros formatos da mesma família (HTML ou um formato específico em XML), ou de natureza diferente: através duma XSLT (*Extensible Style Language Transformation*) são definidas regras para conversão que permitem assegurar a portabilidade dos dados. Um exemplo, ainda exploratório, mas com particular interesse na nossa área são as experiências de conversão de formatos MARC em XML.[xliv]

Nesta introdução simplificada do que é o XML [xlv] podemos desde já apreciar algumas das potenciais vantagens desta tecnologia para ajudar a ultrapassar vários dos problemas da Web [xlvii], entre eles melhorar a qualidade da indexação automática feita por motores de busca, apoiar o processamento automático para outros fins, nomeadamente em operações de gestão de dados e documentos e ainda minorar as questões de preservação de recursos digitais. Neste aspecto, é muito importante o facto de o XML consistir numa linguagem/formato não proprietário, auto-explicativo, e independente de aplicações específicas, o que não acontece, por exemplo, com a maior parte dos formatos de texto que hoje utilizamos (Word, Pdf, etc.).

Um outro aspecto é o que se relaciona com a ligação ao universo das bases de dados. A existência de uma linguagem para Web capaz de representar estruturas de dados específicas e complexas tende a esbater as fronteiras entre esses dois mundos e a tornar possível uma maior convergência técnica o que só terá vantagens futuras em termos de gestão e exploração de informação.[xlviii] Por exemplo, um aspecto fundamental da tecnologia das bases de dados é a existência de linguagens de interrogação *standard*, como é o caso do SQL. Também para o universo de informação estruturada em XML estão já a ser desenvolvidas linguagens de interrogação como o XML Query [xlix].

### **Metadados – informação de referência e controlo nos próprios documentos**

Outra das áreas a concentrar actualmente muitas atenções e actividade é a dos modelos e formatos de metadados, sobretudo metadados descritivos [I]. A definição mais comum para metadados é a de que se trata de “dados sobre outros dados” – uma definição pobre e vaga, quando o que se pretende significar, na maior parte das vezes, é um conjunto de *elementos de informação*, identificativa e descritiva dos recursos a que respeitam.[ii] Metadados descritivos já é uma designação mais consentânea com o conteúdo a que se referem, o qual não é, em conceito e objectivo, nada de muito novo. As descrições bibliográficas e arquivísticas são essencialmente o mesmo. No entanto, o contexto e objectivo funcional destes novos metadados é diferente nalguns aspectos, sendo nisso que se pode justificar a existência de nova terminologia.

Genericamente, um objectivo fundamental dos metadados é a de acrescentar ao próprio recurso electrónico dados que viabilizem maior funcionalidade na interacção com agentes, humanos ou automatizados. Este é o aspecto essencial, quer se trate de metadados descritivos, normalmente vocacionados para a pesquisa e recuperação [lii], quer se trate de outros tipos de metadados, para

efeitos de controlo de direitos e utilização, ou para fins de gestão e preservação, nestes casos designados genericamente de metadados administrativos ou técnicos, respectivamente.

No caso dos metadados descritivos - e contrariamente à tónica que muitas vezes tem sido colocada a propósito de uma alegada “repetição de pior qualidade” - as principais diferenças, face à catalogação que se pratica em bibliotecas e às descrições de arquivo, não residem nem nos elementos de informação a contemplar, nem na sua forma de organização. Esses são aspectos secundários, embora sejam variáveis e discutíveis, face à diferença essencial que reside no contexto em que são gerados e no seu objectivo funcional [liii]; para além de constituírem dados legíveis por computador, os metadados descritivos são informação que se destina a ser:

- parte integrante do próprio recurso, desde a sua origem;
- fornecida pelo autor, ou qualquer outro que seja o agente que disponibilize o recurso em rede;
- a ser não só “lida” mas “compreendida” por agentes automatizados, de forma a poder ser usada para diversos efeitos, nomeadamente de descoberta, selecção, pesquisa e recuperação, e gestão do recurso.

Estes três aspectos visam cumprir funções de controlo que são sincrónicas com o acto da publicação em rede, tecnicamente articuladas com a natureza própria da Web, e contextualmente dirigidas ao todo que é o espaço de informação WWW. Em contrapartida, as funções dos serviços de informação como as bibliotecas são principalmente diacrónicas, são selectivas e são dirigidas a um contexto próprio que não abrange nem toda a Web, nem só a Web.

Teoricamente, a utilidade dos metadados descritivos fornecidos com os próprios documentos é indiscutível. Na prática ainda há um enorme caminho a percorrer até que exista massa crítica para que os efeitos da sua aplicação sejam apreciáveis, a nível de uma comunidade ou a nível geral da WWW. As iniciativas para o desenvolvimento de metadados associados aos recursos electrónicos iniciaram-se em 1995, visando questões para as quais a Web não dispunha de soluções funcionais, em termos de mecanismos lógicos de acesso à informação. Até hoje permanece improvável a ideia de que um único modelo, uma única solução, seja capaz de responder a todas as necessidades.

Na prática, assistir-se-á a uma proliferação – que já é visível – de modelos de metadados para comunidades e serviços de informação diferentes que podem ou não envolver-se com os serviços tradicionais de biblioteca. Entre eles destacar-se-ão alguns, como já acontece com o modelo Dublin Core [liv] que, sendo de aplicação geral, servirão também de modelo e formato de troca para garantir a compatibilidade e interoperabilidade entre diversos tipos de metadados.

O modelo Dublin Core é um modelo de 15 elementos base que pretende ser: fácil de utilizar por quem quer que seja que disponibilize recursos na Internet; flexível, uma vez que os seus elementos não são prescritivos; e expansível, através da qualificação dos elementos base, para necessidades de registo de informação mais detalhada.

Neste sentido, o Dublin Core é passível de desenvolvimento para comunidades e sistemas específicos – estabelecimento de perfis - sem perder, simultaneamente, a compatibilidade com a versão base, o que assegura as condições para um mínimo de interoperabilidade a nível global da WWW. [Iv]

As condições para a interoperabilidade são actualmente mais sensíveis do que nunca já que para resolver problemas de heterogeneidade e especificidade a tendência natural é para uma multiplicidade de normas e especificações. Porque existe consciência dessa situação desenvolvem-se simultaneamente esquemas de mapeamento (*crosswalks*) [Ivi] entre formatos de diferentes modelos de metadados e registos públicos (*metadata registries*) [Ivii] desses mesmos formatos.

O mapeamento entre formatos é hoje uma preocupação básica inerente a qualquer iniciativa relacionada com metadados. A elaboração de esquemas de mapeamento constitui uma oportunidade para identificar questões de portabilidade de dados e para procurar soluções de compatibilidade ou compromisso onde a conversão não é directa. De notar que esta actividade, bem conhecida no âmbito das bases de dados, não se confina actualmente à relação entre novos formatos de metadados mas abrange também formatos tradicionais como os formatos MARC, o que tem subjacente a necessidade de garantir também a interoperabilidade entre os modelos de dados já usados em sistemas de gestão de informação e os modelos de dados a implementar nos próprios objectos de informação. [Iviii]

Por sua vez, os *registos de metadados* destinam-se a declarar as sintaxes e semânticas de diferentes modelos e formatos de metadados, tornando publicamente disponíveis as suas descrições, os seus objectivos, estruturas, regras de aplicação, entidades responsáveis, relações com outros modelos de metadados, etc. Os registos de metadados começam agora a surgir e terão um importante papel na promoção da interoperabilidade, contribuindo para disseminar as especificações existentes, evitar a duplicação de esforços e promover a relação entre comunidades usando diferentes metadados. Esta actividade de registo público é, aliás, enquadrada a nível da ISO (ISO 11174 [Iix] e BSR [Ix]) pois compreende uma necessidade global que não é só da modelação de dados, mas também da semântica básica em geral usada em sistemas de gestão de informação.

#### **RDF – uma arquitectura independente para metadados**

Enquanto “informação de referência” os metadados descritivos podem apresentar questões complexas, mas não muito diferentes daquelas a que estamos habituados, nos sistemas de informação bibliográfica. Já do ponto de vista de representação as coisas são um pouco diferentes, uma vez que normalmente esses metadados, destinando-se à Web, devem ser expressos em HTML ou XML. Originalmente os metadados, nomeadamente no formato Dublin Core, destinavam-se a ser expressos em HTML, usando o elemento META, a partir das versões 2.0 e superiores. Na especificação do HTML 4.0 foram incluídos atributos adicionais para o elemento META, de forma a permitir implementações mais específicas. No entanto, subsistem limitações quanto à coexistência de

diferentes formatos de metadados, relações entre elementos, suas fontes, etc., para a sua cabal interpretação e uso, nomeadamente por programas de computador.

Por outro lado, também a simples inclusão dos metadados numa estrutura geral, que é variável, em XML, não resolve todos os requisitos específicos dos metadados, com a eficiência desejada. O RDF – *Resource Description Framework* [Ixii], surgiu para responder a essas necessidades. É uma aplicação específica do XML, constituída por um modelo e uma sintaxe genéricos que permitem a acomodação de metadados de uma forma que é independente da estrutura e semântica próprias desses metadados. [Ixiii] Trata-se de um mecanismo neutro que simplifica a arquitectura do recurso electrónico, criando um “contentor” próprio para os metadados, distinguindo-os claramente do conteúdo do documento em si; ao mesmo tempo, permite a extensibilidade necessária à formulação de relações complexas entre os elementos da descrição e destes com fontes exteriores que autorizam ou dão informação sobre esses elementos ou suas propriedades. [Ixiii]

Neste sentido, o RDF não só facilita o transporte de qualquer tipo de metadados, funcionalmente um pouco à semelhança do ISO 2709 para qualquer tipo de formato MARC, possibilitando a coexistência de diversos formatos de metadados, como permite ainda eliminar redundâncias, facilitar o processamento automatizado dos dados que constituem os metadados e a interoperabilidade. O RDF aplica-se a qualquer tipo de metadados, para qualquer finalidade, sendo particularmente promissor no que respeita aos metadados descritivos.

XML e RDF são, assim, duas peças fundamentais à engrenagem da interoperabilidade semântica necessária para conferir valor acrescentado à amálgama de recursos Web pouco estruturados, dificilmente geríveis e com dados pouco partilháveis por diferentes sistemas e aplicações. São dois elementos técnicos fundamentais para operacionalizar a Web semântica mas, por si só, não resolvem as questões de adequação e qualidade de conteúdos. A adequação e qualidade de conteúdos dos metadados dependerão, essencialmente, das políticas, dos requisitos, das metodologias e das práticas que forem sendo construídas pelas diferentes comunidades face às funcionalidades por elas pretendidas. [Ixiv]

Duma forma geral, quer o aparecimento de qualificadores [Ixv] dentro do próprio modelo Dublin Core quer a especificação de desenvolvimentos mais detalhados e de critérios de implementação para comunidades e finalidades específicas [Ixvi] torna evidente que as questões semânticas não são fáceis, invadem a modelação de dados mas não se resumem a ela, e que um modelo de dados simplista e não apoiado por outros factores, intelectuais e organizacionais, não suporta senão funcionalidades simples e apenas responde a requisitos de interoperabilidade também simples. O que nem sempre é suficiente. É aqui que é pertinente recordar a experiência e lições do Z39.50, não esquecendo que tem atrás de si toda uma experiência assente em mecanismos, princípios, normas e regras para controlo de objectos de informação, adoptados por uma comunidade que é alargada, mas não deixa de ser uma comunidade, e onde, mesmo assim, não é fácil resolver todas as questões.

O equilíbrio entre os requisitos de qualidade e funcionalidade da informação de controlo – que são os metadados – e a sua praticabilidade a nível geral da WWW, é difícil de estabelecer já que a Internet não é um sistema, como o são uma biblioteca, um *gateway* de informação, ou mesmo uma biblioteca digital, distribuída ou não. A Internet é uma realidade aberta e não controlada, em termos de gestão de recursos e de funções, em que a heterogeneidade se sobrepõe, por natureza, à normalização.

### CONVERGÊNCIA DE CULTURAS TECNOLÓGICAS – QUE NOVOS PARADIGMAS?

Não se aplicam, assim, à Internet as premissas próprias da natureza de um *sistema* – enquanto realidade delimitada, composta por elementos internos que se relacionam entre si de forma arquitectada, em função de objectivos pré-estabelecidos, relacionando-se como um todo com o seu ambiente exterior e evoluindo na interacção com ele. A Internet, e principalmente a WWW, podem ser vistos ao mesmo tempo como realidades interiores e exteriores, dependendo da escala e ponto de vista que se toma para olhar o espaço informacional, ou partes dele. Não é por acaso que as noções de *ambiente* e de *espaço* estão tão associadas aos conceitos de Internet e WWW. Mas pode a Internet, ou a WWW, enquanto realidade técnica e enquanto conteúdos ser apenas o ambiente ou o espaço de si própria?

Com a necessidade de dotar a WWW de mecanismos que a tornem mais “gerível”, o que está a acontecer é a introdução de elementos característicos do mundo dos sistemas suportados por bases de dados – ontologias, modelação de dados, relações lógicas, arquitecturas de informação, interoperabilidade, linguagens de interrogação de dados estruturados – no mundo dos standards que se pretendem básicos para a Web do futuro, como o são o XML e os metadados. Do mesmo modo que os sistemas de bases de dados se modificam nos seus paradigmas para arquitecturas por níveis e distribuídas, em que a relação e interdependência com recursos e tecnologias que lhe são externas também aumenta. Estas transformações começam a esbater as fronteiras entre esses dois mundos, modificando os respectivos paradigmas.

Z39.50, XML, metadados como o Dublin Core e RDF são exemplos de tecnologias cuja filosofia é indicativa das tendências actuais em matéria de produção, gestão e utilização de recursos e serviços de informação. Não significa isto que qualquer destas soluções tecnológicas seja única no seu campo. De facto, quase todas elas, inclusive o tão recente RDF, têm alternativas mais ou menos equivalentes, nascidas mais ou menos nas mesmas alturas, mas em comunidades diferentes. O facto de serem estas as referidas é porque estarão, pelo contexto em que apareceram e que as suporta – entidades como a NISO/ISO/Library of Congress, OCLC, e Consortium 3W – em maior destaque e em melhor posição para lograr os seus objectivos.

Mas também não é tão importante assim saber prever se serão exactamente estas – tal como agora se apresentam – as soluções que vingarão no futuro. O importante é

perceber se têm subjacente uma filosofia com futuro, e qual é, para as questões actuais. Se essa filosofia existir e for consistente e sustentável, ela consolidar-se-á, com as actuais soluções tecnológicas ou com outras que surjam e que porventura melhor a realizem.

Não se trata, por isso, de ter que se conhecer em detalhe toda a parafernália de novas normas e especificações técnicas que nos são propostas, que passam à prática e ficam, ou que se sucedem e substituem, a uma cadência dificilmente digerível. Em certa medida, e relativamente a muitos pormenores técnicos, uma dieta inteligente é mesmo melhor.

A evolução do mundo da informação não é hoje separável das transformações da tecnologia. Conhecer o significado e a filosofia geral dessa evolução é importante para antecipar, e ajudar a formular, conceitos chave com os quais é importante contar na construção de novos modelos para serviços de intermediação de informação, como são as bibliotecas.

Como irão evoluir os seus repositórios de informação e mecanismos de gestão, e os serviços que prestam, a partir do que são hoje funções centralizadas e independentes para um ambiente de funções distribuídas e interdependentes amplamente promovido pelas tecnologias? Por exemplo, que alcance, que modelo e que modo(s) de realização prática sustentável terá a função de “controlo bibliográfico” face à natureza das questões e à escala do problema que a informação disponível em rede representa? [Ixvii]

É no sentido destas questões que é importante, no momento actual, posicionarmo-nos com um bom ângulo de visão sobre a imagem global do que se passa na cena tecnológica. Com a preocupação principal de não deixar escapar aqueles aspectos que, no futuro, podem fazer toda a diferença.

---

### NOTAS

<sup>i</sup> Estruturalmente, a situação ainda é hoje, por exemplo, muito idêntica à que Michael Lesk descrevia em *Mad Library Disease: Holes in the Stacks*, em 1996. [Em linha]. URL: <http://www.lesk.com/mlesk/ucla/ucla.html>

<sup>ii</sup> RUSBRIDGE, Chris - *Towards the hybrid library*. *D-Lib Magazine*. [Em linha]. July/August 1998. URL: <http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html>; ver também uma recolha de opiniões sobre esta questão em OPPENHEIM, Charles; Smithson, Daniel. *What is the hybrid library?*. *Journal of Information Science*, 25 (2) 1999, p.97-112.

---

<sup>iii</sup> COLLIER, Mel - Towards a general theory of the Digital Library. [Em linha]. URL: <http://www.dl.ulis.ac.jp/ISDL97/proceedings/collier.html>

KNIGHT, J. - The Hybrid Library: Books and Bytes. Ariadne. [Em linha]. 11 (1998). URL: <http://www.ariadne.ac.uk/issue11/knight/>

<sup>iv</sup> BORGMAN, Christine - Is it Digital or is it a Library? Digital libraries and Information Infrastructure. In Borgman, C. - From Guttenberg to the Global Information Infrastructure. Cambridge, Mass.; London : the MIT Press, 1999, p.34-52.

<sup>v</sup> Ver MURRAY, I. R. - Assessing the effect of new generation library management systems. *Program*, vol.31,4 (1997) p.313-327.; e AKEROYD, John; COX, Andrew. Integrated Library Mangement Systems: an Overview. *VINE*, 115 (1999) p. 3-10.

<sup>vi</sup> Ver, a título de exemplo, uma amostragem de características gerais em The ILSR Vendor Survey 1999: Full Results. [Em linha]. <http://www.ilsr.com/SURVEY2.HTM>

<sup>vii</sup> DZURINKO, Mary K. Integrated Online Library Systems. (last rev. June 1998). ILSRT. [Em linha]. URL: <http://www.ilsr.com/iols.htm>

<sup>viii</sup> MYBURGH, Sue – The convergence of information technology & information management. *The Information Management Journal*, April 2000, p.4-16.

<sup>ix</sup> GARROD, Penny – Survival strategies in the Learning Age – hybrid staff and hybrid libraries. *ASLIB Proceedings*, vol. 51. No 6 (June 1999), p.187-194.

<sup>x</sup> Na Europa, esta fase foi marcada pelos programas promovidos pela CE desde 1991, que , na sequência da Resolução de 17 Set. 1985 (JOCE NC 271, de 1985.10.23 – “Collaboration between libraries in the field of data processing”) incluíram as bibliotecas como uma das áreas do Programa *Telematics Applications*, continuado pelo Programa IST. URL: <http://www.cordis.lu/ist/home.html>

<sup>xi</sup> DAVIES, Clare; Scammell, Alison; Hall, Matthew - eLib: Changing the Lightbulb - er, the Culture. *Ariadne*. [Em linha]. 10 (1998). [Em linha]. URL: <http://www.ariadne.ac.uk/issue10/cultural/>

<sup>xii</sup> Sobre o impacto dos programas europeus ver: STORK, Hans-Georg - Ten Years of European Commission Support to European Libraries: Results and Perspectives. Nov. 98. [Em linha]. URL: <http://158.169.50.95:10080/libraries/en/fribourg.html>  
e  
European Commission. DG Information Society. Impact of the Telematics for Libraries Programme under the Fourth Framework Programme. 1999. [Em linha]. URL: <ftp://ftp.cordis.lu/pub/ist/docs/digicult/impact.pdf>

---

<sup>xiii</sup> MCCALLUM, Sally - What makes a standard ? Cataloguing and Classification Quarterly, vol.21, 3-4 (1996) p.5-15.

<sup>xiv</sup> Ver, por exemplo uma comparação dos processos de produção de normas entre ISO/IEC e o IETF em PALME, Jacob – Notes on Standards Making. (Up. 5 July 1995). [Em linha]. URL: <http://www.dsv.su.se/jpalme/standards-research.html>

<sup>xv</sup> CAPLAN, Priscilla - Oh What a Tangled Web We Weave: Opportunities and Challenges for Standards Development in the Digital Library Arena. *First Monday*, vol. 5, No 6 - June 5th 2000. [Em linha].URL: [http://firstmonday.org/issues/issue5\\_6/caplan/](http://firstmonday.org/issues/issue5_6/caplan/)

<sup>xvi</sup> MCLEAN, Neil – Creating Information Futures: the quest for sustainability. *Proceed. of Information online & on disc 99*, Sydney, 1999. [Em linha]. URL: <http://www.csu.edu.au/special/online99/proceedings99/310.htm>

<sup>xvii</sup> BARRY, Tony – The next waves of change: the future as seen from January 1999. *Proceed. of Information online & on disc 99*, Sydney, 1999. [Em linha]. URL: <http://www.csu.edu.au/special/online99/proceedings99/210.htm>

<sup>xviii</sup> Microsoft Corporation – Why is interoperability important? Microsoft Technet. [Em linha]. URL: <http://www.microsoft.com/technet/interop/ndam.asp>

<sup>xix</sup> MILLER, Paul - Interoperability: What is it and Why should I want it? Ariadne. [Em linha]. 24 (2000). URL: <http://www.ariadne.ac.uk/issue24/interoperability/>

<sup>xx</sup> Ver, a título de exemplo UK Interoperability Focus, estabelecido no início de 1999. [Em linha]. URL: <http://www.ukoln.ac.uk/interop-focus/about/>

<sup>xxi</sup> NELL, J. G. – Designing standards to improve process interoperability. (March 2000). [Em linha]. URL: <http://www.mel.nist.gov/sc5wg1/wg1vision.htm>

<sup>xxii</sup> ISO/IEC 7498-1:1994 - Information technology - Open Systems Interconnection -- Basic Reference Model: The Basic Model

<sup>xxiii</sup> Para informação geral ver International Standard Z39.50 Maintenance Agency. URL: <http://lcweb.loc.gov/z3950/agency/> e NISO Z39.50 Resource Page. URL: <http://www.niso.org/z3950.html>

<sup>xxiv</sup> LOPES, Maria Inês – A importância da interoperabilidade entre sistemas de informação bibliográfica : a norma Z39.50. Com.ao 6º Congresso BAD, Aveiro, 1998.

MILLER, Paul - Z39.50 for All. *Ariadne*. [Em linha]

(Sept. 1999). URL:  
<http://www.ariadne.ac.uk/issue21/z3950/intro.html>

<sup>xxv</sup> Por exemplo, diferenças no número e tipo dos índices de pesquisa disponíveis, nas funcionalidades sobre as chaves de pesquisa – como tipo(s) de truncatura, posição, relação, etc. - nos *sets* de caracteres suportados, nas regras de alfabetação implementadas, etc.

HINNEBUSH, Mark – Z39.50 : Report to the CIC on the State of Z39.50 Within the Consortium. Final rev. 29 Aug. 1999. [Em linha]. URL:  
<http://www.cic.uiuc.edu/cli/z39-50report.htm>

<sup>xxvi</sup> DEKKERS, Makx - Implementing Z39.50 in a Multi-national and Multi-lingual environment. (Cop. 1995). [Em linha]. URL:  
<http://lcweb.loc.gov/z3950/agency/abstracts/makx.html>

LUNAU, Carrol; Turner, Fay - Summary of Issues Related to the Use of Z39.50. (Rev. 1997.07.18). [Em linha]. URL:  
<http://www.nlc-bnc.ca/resource/vcuc/ezarlsum.htm>

<sup>xxvii</sup> The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery. Release 1.1. (Up. Feb. 2001). [Em linha]. URL:  
<http://www.nlc-bnc.ca/bath/bp-current.htm>

<sup>xxviii</sup> MOEN, William – Resource Discovery Using Z39.50: Promise and Reality. (Nov.2000). [Em linha]. URL:  
<http://lcweb.loc.gov/catdir/bibcontrol/moen.html>

<sup>xxix</sup> Nomeadamente, conforme refere, Le Van, por ser um protocolo *stateful* e usar codificação binária, o que não se enquadra nos pressupostos da Web.  
LE VAN, Raph – Library Experience in Online Searching. A position paper for the Query Languages Worksop, 1998. [Em linha]. URL:  
<http://www.w3.org/TandS/OL/OL98/pp/libraryexp.html>

<sup>xxx</sup> TAYLOR, Mike – Z39.50 – The Empire Strikes Back. (6 Dec. 2000). [Em linha]. URL:  
<http://www.miketaylor.org.uk/tech/esb.html>

<sup>xxxi</sup> GATENBY, Janifer – Internet, Interoperability and Standards: Filling the Gaps. (Mod. Aug. 2000). [Em linha]. URL: <http://www.niso.org/Gatenby.html>

<sup>xxxii</sup> ZIG – Dec. 2000 Meeting Output. [Em linha]. URL:  
<http://www.loc.gov/z3950/agency/zig/meetings/dc2000/output.html>

<sup>xxxiii</sup> Quanto mais os serviços de informação são eficazes e fáceis de usar mais escondem a complexidade e sofisticação das competências profissionais subjacentes...  
KLING, Rob - Information Technologies and the Strategic Reconfiguration of Libraries in Communication Networks. (Draft 1.0, 12/5/2000). [Em linha]. URL:  
<http://www.slis.indiana.edu/csi/wp00-04.html>

<sup>xxxiv</sup> MELNIK, Sergey; DECKER, Stefan – A layered Approach to Information Modeling and Interoperability

on the Web. (Rev. Sep. 4, 2000). [Em linha]. URL:  
<http://www-db.stanford.edu/~melnik/pub/sw00/>

<sup>xxxv</sup> OJALA, Marydee – The end of online as we know it. Proceed. of Information online & on disc 99, Sydney, 1999. [Em linha]. URL:  
<http://www.csu.edu.au/special/online99/proceedings99/100.htm>

<sup>xxxvi</sup> Este aspecto é particularmente abordado em  
ABITEBOUL, S. ; BUNEMAN, P; SUCIU, Dan – Data on the Web : From Relations to Semistructured Data and XML. San Francisco, CA: Morgan Kaufmann, 2000. ISBN 1-55860-622-X.

<sup>xxxvii</sup> Ver exemplos de arquitecturas *broker* em  
PROJECT RENARDUS. Evaluation Report of existing broker models in related projects. (31 March 2000) . 90p.  
[http://www.renardus.org/deliverables/d1\\_1/D1\\_1\\_final.pdf](http://www.renardus.org/deliverables/d1_1/D1_1_final.pdf)

<sup>xxxviii</sup> NOTESS, Greg - Notess.Com. Search Engine Showdown . [Em linha]. URL:  
<http://searchengineshowdown.com/>

<sup>xxxix</sup> LEE, T. Berners – Web architecture from 50,000 feet. (Up.Oct 1999). [Em linha]. URL:  
<http://www.w3.org/DesignIssues/Architecture.html>

<sup>xl</sup> W3Consortium – Semantic Web Activity Statement. [Em linha]. URL: <http://www.w3.org/2001/sw/Activity>

<sup>xli</sup> SWAG (Semantic Web Agreement Group) – What is the Semantic Web? . [Em linha]. URL:  
<http://swag.semanticweb.org/whatIsSW>

<sup>xlii</sup> LEE, T. Berners – Evolvability. (Ed. 200.09.08). [Em linha].URL:  
<http://www.w3.org/DesignIssues/Evolution.html>

<sup>xliii</sup> W3Consortium Extensible Markup Language (XML) 1.0 (Second Edition). (6 Oct. 2000). [Em linha]. URL:  
<http://www.w3.org/TR/REC-xml>.

<sup>xliv</sup> CONNOLY, Dan; KHARE, Rohit; RIFKIN, Adam – The Evolution of Web Documents: The Ascent of XML.(98.01.15) [Em linha]. URL:  
<http://www.cs.caltech.edu/~adam/papers/xml/ascent-of-xml.html>

<sup>xlv</sup> CROSSNET SYSTEMS LTD – Using XSLT For XML MARC Record Conversion. (Discus. paper, v. 0.2, 16.06.2000). [Em linha]. URL:  
[http://roadrunner.crxnet.com/one2/xslt\\_marc\\_report.pdf](http://roadrunner.crxnet.com/one2/xslt_marc_report.pdf)

<sup>xlvi</sup> Para uma boa selecção de recursos sobre XML ver

---

COVER, Robin – The XML Cover Pages. [Em linha]. URL: <http://xml.coverpages.org/sgml-xml.html>

<sup>xlvii</sup> BOSAC, Jon – XML, Java and the Future of the Web. (Rev. 1997.03.10). <http://www.ibiblio.org/pub/sun-info/standards/xml/why/xmlapps.htm>

<sup>xlviii</sup> BOURRET, Ronald - XML and databases. (Up. Nov. 2000). [Em linha]. URL: <http://www.rpbouret.com/xml/XMLAndDatabases.htm>

<sup>xlix</sup> W3Consortium - XML QUERY. (Rev.2001.02.16) <http://www.w3.org/XML/Query>

<sup>l</sup> BACA, Murtha, ed. – Introduction to Metadata: Pathways to Digital Information. The J. Paul Getty Trust, 1998. [Em linha]. URL: <http://www.getty.edu/research/institute/standards/intrometadata/index.html>

<sup>li</sup> DOVEY, M.J. – “Stuff” about “Stuff” – the different meanings of metadata. *VINE*, 116 (2000), p.6-13.

<sup>lii</sup> RUSH-FEJA, Diann – Metadata: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources). 1998. [Em linha]. URL: <http://www.eso.org/gen-fac/libraries/lisa3/ruschfejad.html>

<sup>liii</sup> LEE, T. Berners – Metadata Architecture. (last ed. 2000.09.21). [Em linha]. URL: <http://www.w3.org/DesignIssues/Metadata>

<sup>liv</sup> Dublin Core Metadata Element Set, Version 1.1: Reference Description. (1999.07.02). [Em linha]. URL: <http://dublincore.org/documents/dces/>

<sup>lv</sup> WEIBEL, Stuart - The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions. *D-Lib Magazine*. [Em linha]. Vol. 6, 12 (Dec. 2000). URL: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>

<sup>lvi</sup> DAY, Michael – Metadata: Mapping between metadata formats. (Up. 13.02.2001). [Em linha]. URL: <http://www.ukoln.ac.uk/metadata/interoperability/>

<sup>lvii</sup> DESIRE Metadata Registry Framework. (31.03.2000). [Em linha]. URL: <http://www.desire.org/html/research/deliverables/D3.5/d35.html>

<sup>lviii</sup> Library of Congress. MARC SGML and XML. MARC DTDs. [Em linha]. URL: <http://lcweb.loc.gov/marc/marcsgml.html>

MILLER, Dick R. - XML: Libraries' strategic opportunity. *Library Journal NetConnect*. [Em linha].

---

Sum.2000. L:<http://www.libraryjournal.com/xml.asp>

QIN, Jon - Representation and organization of Information in the Web Space: From MARC to XML. *Informing Science*. [Em linha]. URL: <http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>

<sup>lix</sup> ISO/IEC 11179-1 Specification and standardization of data elements. Final Draft International Standard. [Em linha]. URL: <http://www.sdct.itl.nist.gov/~ftp/18/11179/>

<sup>lx</sup> ISO TC 154 "Processes, data elements and documents in commerce, industry and administration". WG1 "Basic Semantics Register" (BSR). [Em linha]. URL: <http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/TC154WG1/index.htm>

<sup>lxi</sup> Resource Description Framework. (RDF) Schema Specification 1.0. ( 27 march 2000). [Em linha]. URL: <http://www.w3.org/tr/2000/cr-rdf-schema-20000327/#refs>

<sup>lxii</sup> LEE, T. Berners - Why RDF model is different from the XML model. [Em linha]. URL: <http://www.w3.org/DesignIssues/RDF-XML.html>

<sup>lxiii</sup> BRICKLEY, Dan -Understanding RDF. (v.2001.02.22). [Em linha]. URL: <http://ilrt.org/discovery/2001/01/understanding-rdf/>

<sup>lxiv</sup> DILLON, Martin – Metadata for Web resources: how metadata works on the Web. *Bicentennial Conf. on Bibliographic Control for the New Millemium, LoC, 2000*. [Em linha]. URL: <http://www.loc.gov/catdir/bibcontrol/dillon.html>

<sup>lxv</sup> Dublin Core Qualifiers. (2000.07.11). [Em linha]. URL: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

<sup>lxvi</sup> CIMI. Guide to best Practice: Dublin Core. V1.1(21.04.2000). [Em linha]. URL: [http://www.cimi.org/public\\_docs/meta\\_bestprac\\_v1\\_1\\_2\\_10400.pdf](http://www.cimi.org/public_docs/meta_bestprac_v1_1_2_10400.pdf)

<sup>lxvii</sup> LYNCH, Clifford – The New Context for Bibliographic Control in the New Millenium. (Nov. 2000). [Em linha]. URL: <http://www.lcweb.loc.gov/catdir/bibcontrol/lynch.html>